SciDAC Visualization and Analytics Center for Enabling Technologies
Project Management Plan
Version 1.1

E. Wes Bethel [*] and Chris Johnson [†]
Principal Investigators

Charles Hansen, Claudio Silva, Steven Parker, Allen Sanderson,
Xavier Tricoche, Martin Cole [‡]

Sean Ahern, George Ostrouchov, Jeremy Meredith [§]

Valerio Pascucci, Hank Childs, Peer-Timo Bremer,
Daniel Laney, Ajith Mascarenhas, Kathleen Bonnell [¶]

Ken Joy, Bernd Hamann [‖]

May 2007

[*] Lawrence Berkeley National Laboratory
[†] Scientific Computing Institute, University of Utah
[‡] Scientific Computing Institute, University of Utah
[§] Oak Ridge National Laboratory
[¶] Lawrence Livermore National Laboratory
[‖] University of California – Davis

Contents

1 Mission

The SciDAC Visualization and Analytics Center for Enabling Technologies (VACET) focuses on leveraging scientific visualization and analytics software technology as an enabling technology for increasing scientific productivity and insight. Our mission is to foster scientific insight through creating and deploying effective data understanding technology that is truly responsive to the needs of our stakeholders in the scientific research community. Our stakeholders – researchers from from experimental and computational sciences – are "awash in data" produced by an "information big bang" whereby their ability to collect and generate data far out paces the ability to understand such data. This challenge is widely acknowledged to be one of the primary bottlenecks in contemporary science.

The vision for our Center is to respond directly to this challenge by adapting, extending, creating when necessary and deploying visualization and data understanding technologies for our science stakeholders. Using an organizational model Center for Enabling Technologies, we are well positioned to be responsive to the needs of a diverse set of scientific stakeholders in a coordinated fashion using a range of visualization, mathematics, statistics, computer and computational science and data management technologies.

2 Goals and Objectives

SciDAC's stated mission is to use advanced computing to further the DOE research programs in basic energy sciences, biological and environmental research fusion, and high-energy and nuclear physics. More specifically, the SciDAC program aims to produce "a Scientific Computing Software Infrastructure that bridges the gap between advanced research in applied mathematics and computer science, and computational science research in the physical, chemical, biological and environmental sciences [2]." Stated differently, SciDAC's mission is enable advances in scientific understanding through use of high performance computational platforms.

SciDAC will achieve its mission through several interrelated thrusts. First, Science Applications produce computational codes that simulate scientific phenomenon ranging from models of exploding supernovae to predicting function of new protein structures. Second, Centers for Enabling Technology (CETs) produce computational infrastructure that play varying roles in helping Science Applications to achieve their goals. Projects here include core numerical solvers suitable for use on large parallel machines, visual data analysis technology to help Science Applications derive understanding from data produced by their codes to performance models that predict how well a given software application will perform on a given architecture. Third, Partnerships are focused efforts involving personnel from Science Applications, Centers and possibly Institutes (below). These are teams of experts from different areas that combine forces to produce novel solutions. Fourth, Institutes (according to the FY06 call[1] are somewhat more broadly defined: University-lead centers of excellence intended to complement CET's that [...] focus on major software issues through a range of collaborative research interactions. Stated differently, Institutes serve as a focal point for research activities in a given area, reach out to engage of broader activities of scientific discovery, have a dimension of training and outreach in high performance computing topics including graduate students and post-docs.

The SciDAC program as a whole is evaluated by its quality of scientific output. In the case of Science Applications, this means new scientific discoveries. More broadly, it means all elements of

[1] http://www.science.doe.gov/grants/FAPN06-04.html

the SciDAC program are expected to produce top quality research as evidenced by peer-reviewed publications.

As a CET, one of VACET's central goals is to bring cutting-edge, production quality visualization and analytics to bear on the data understanding challenges of the SciDAC scientific community in the form of well engineered, usable software. Achieving this overall mission objective will entail accomplishing several interrelated goals. Achieving these goals is possible only through our organization as a SciDAC Center for Enabling Technologies, as described later in Section 3.

2.1 Foster Effective Stakeholder Relationships

One of the central pillars of our Center is strong relationships with our science stakeholders. The specific goal here is achieving vibrant, productive interactions with our customer/stakeholders. We translate their needs into specific, actionable items that in turn result in new scientific data understanding capabilities.

2.2 Coordinated Interactions with Stakeholders

Like most within the scientific research community, our stakeholders' time is scarce and valuable. Truly effective data understanding solutions emerge only through meaningful and ongoing dialog between stakeholder and technology provider. Our stakeholders typically do not have the luxury of duplicating such interactions across a number of different data understanding providers. One goal of our center, which is possible only through organization as a Center for Enabling Technologies, is to focus stakeholder interactions through a VACET "Customer Project Manager," who is responsible for such interactions between stakeholder and VACET. We have organized our center as described in Section 6, to optimize the all-important stakeholder-VACET relationship, which is a crucial element of success to both stakeholder and VACET.

2.3 Effective Production Deployment of Petascale-Capable Technology

One of VACET's goals is to deploy its production-quality, petascale-capable data understanding technology at DOE's Open Computing facilities. To that end, VACET team members include team leads from visualization and analytics at the National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory and the Leadership Computing Facility (LCF) at Oak Ridge National Laboratory. A motivation for such a focus is the fact that most of DOE's computational science activities are conducted at these facilities. We anticipate in the future an increasing trend towards consolidation of community-wide data repositories for computational and experimental sciences at these facilities. Therefore, we have engineered into our team's organization and makeup close relationships with production visualization and analytics activities at these DOE computing facilities.

2.4 Coordinated Interactions with Other SciDAC Technology Providers

One of the strengths of the SciDAC model is the opportunity for leveraging multiple technologies to provide focused solutions. In this spirit, VACET aims to leverage as much existing technology from other SciDAC technology providers as appropriate to provide integrated solutions to our stakeholders by reducing our development time and accelerating our time to solution. Activities towards achieving this goal strengthen not only VACET but other individual SciDAC efforts as well as the overall SciDAC program.

2.5 Technical Excellence, Publications, Outreach, Education and Service

We anticipate that most of VACET's accomplishments will be groundbreaking, "first-ever" works that span a diverse range of research, development, engineering, deployment, and novel application to challenging data understanding problems. As a consequence, our team feels it important to achieve a high level technical excellence and correspondingly a leadership role in the field of high performance visual data understanding. As communication of results is an important element of our work, we expect our team members to be high achievers in terms of technical publications to convey our accomplishments and the success of the SciDAC program.

Outreach and education activities help to increase the "surface area" between SciDAC, VACET and the scientific community as well as the general public. Outreach activities include things like invited talks, workshops and tutorials. Education activities include opportunities for students to engage in work on "real world" problems, to be exposed to a highly productive, diverse and dynamic team comprised of leaders in the field of visual data understanding. Service includes activites like conference chair, program committee member, and technical paper/proposal reviewer for journals, conferences and DOE programs.

2.6 Train the Next Generation of Visualization and Analytics Leaders

Related, we feel an important VACET goal is to train the next generation of leaders in the field of high performance visual data understanding. In addition to the opportunities for young student scientists, our Center is organized (see Section 6) in such a way as to provide a great deal of opportunity for interaction with stakeholders, other VACET team members, and, perhaps most importantly, opportunity for leadership roles within the Center in the form of Customer Project Manager and technical project Team Leader. All VACET team members participate in project management activities, and all are expected to generate technical publications (papers, web pages, etc.) and to prepare and deliver presentations.

2.7 Provide Positive Return on SciDAC Investment

Since this effort is the first major coordinated visualization activity in DOE in recent years, all on our team are committed to ensure that DOE's investment pays handsome dividends in the form of positive and significant scientific impact.

3 Benefits

SciDAC's stated mission is to use advanced computing to further the DOE research programs in basic energy sciences, biological and environmental research fusion, and high-energy and nuclear physics. More specifically, the SciDAC program aims to produce "a Scientific Computing Software Infrastructure that bridges the gap between advanced research in applied mathematics and computer science, and computational science research in the physical, chemical, biological and environmental sciences." [2].

The first five years of the SciDAC program (2001-2006) have produced dramatic advances in scientific software technology that can take advantage of the world's most powerful computing systems. While most of the SciDAC investment portfolio has focused on scientific software and computing, a small fraction has been dedicated to management of the resulting data. In the first round of SciDAC, there was no organized or coordinated data understanding or visualization effort. The result is a scientific community that has the ability to generate vast amounts of data

and some ability to store and retrieve data, but only a minimally coordinated, effective, deployed, and supported set of software tools and techniques for data understanding.

While the general benefits of scientific visualization are generally well understood and accepted, there are a number of unique benefits that result from our team organized and operated as an integrated Center for Enabling Technologies within the SciDAC program.

3.1 Integrated, Coordinated Visualization and Analytics R&D Activities

Visualization and analytics activities within DOE have traditionally been conducted as small, uncoordinated efforts. With a very few isolated exceptions, most notably the concerted visualization effort within the ASC program [1], there has been no impetus for a DOE-funded visualization scientist to go beyond publishing a research paper, making a pretty picture, and occasionally producing a stand-alone application. The approach we take here is to consolidate and coordinate efforts of leading visualization researchers to support the open science community. Our mission involves the coordinated activities of several different groups and thus requires long term funding to enable us to foster and maintain relationships with application scientists. Such relationships ensure that visualization solutions are responsive to stakeholders' needs, are deployed in the appropriate places, and are supported over the course of a given science project. This type of approach, and the resulting benefit, is possible only through organization as a Center; such results are not possible at the scale of the individual project.

3.2 Leveraging Other SciDAC Technologies to Accelerate Progress

As a Center, we can coordinate the development and deployment activities of related and complementary technologies such as data management, analysis, and visualization. Further, as technologies are developed for a specific application we will be in a position to query other applications areas more broadly deploy them. This type of coordinated effort is not possible at the small project scale of the individual PI. As a Center, we can serve a greater number of stakeholders than would be possible if organized as a number of smaller, isolated efforts.

3.3 Integrated and Effective Deployment

As a Center, we are in a position to interact with other technology centers – Mathematics and Computer Science – to ensure that our data understanding technology is deployed as part of a larger technology ecosystem. Again, that level of broad interaction is not possible at the level of the individual PI. Related, our team includes leaders from production visualization and analytics efforts at NERSC/LBNL and LCF/ORNL, thereby maximizing opportunity for successful deployment at those facilities, which are host to many SciDAC and other computational and experimental sciences.

3.4 Effective Solutions from a Broad Technology Base

Our experience has shown that there simply is no "single visualization technology solution" (no "one size fits all" solution) that is responsive to the broad set of challenges facing the scientific research community. Instead, effective solutions require the careful adaptation and deployment of technologies from many sources. Some technologies, such as application frameworks, are general purpose, while other technologies, like visual presentation paradigms showing the metabolic pathways within a bacterium, are uniquely specific to a given application area. The ability to vet a broad technology base within a Center helps ensure effective solutions and also protects DOE's investment in those constituent technologies.

3.5 Enable Scientific Discovery

With the creation, blending, application and deployment of technology, we are in a unique position to provide markedly new scientific data understanding capabilities. Our operational model, which relies on substantive and long-term relationships with stakeholder/customers, ensures our work focuses on addressing scientific data understanding needs, is engineered to deliver scientific impact.

3.6 Time and Cost Savings

In the past, some of our customers have been forced to allocate a portion of their scarce science research budget to create visualization and analytics infrastructure. Unanimously, these stakeholders desire to get out of the business of developing and maintaining visualization infrastructure. One aspect of our strategy is to provide community-wide visualization infrastructure that will thereby result in cost and time savings to scientific applications: time and effort previously spent on building and maintaining visualization infrastructure can instead be expended on science rather than software engineering.

We expect to realize time and cost savings in many other areas as well. For example, our organization as a Center allows us to focus R&D efforts into core technologies that will then be applied to a variety of different customer projects. A long funding horizon allows us to exercise technology options over different time scales from short- to long-term. Such a model – centralized and coordinated R&D that is then applied in "products" reflects successful business models in industry.

4 Problem

Since the advent of computing, the world has experienced an "information big bang", an explosion of data. Information is being created at an exponential rate, such that since 2003, new information generated annually exceeds the information contained in all previously created documents. Furthermore, digital information now makes up more than 90% of all information produced, vastly exceeding that generated on paper and film. One of the greatest scientific and engineering challenges of the 21st century is to effectively understand and make use of this growing wealth of information to scientific advantage. Software plays a central role in providing the means to manage and understand relationships, anomalies, trends, and features contained in today's abundant scientific data.

In the past decade, the art and science of data understanding in general, and scientific visualization in particular, have been subject to the same challenges as all other scientific fields affected by the explosive growth in the size and complexity of data. The visualization community has responded with a diverse range of research spanning the creation of scalable techniques for increasing visualization throughput, the development of techniques for multivariate visual display and exploration, and a host of other activities [14, 8, 24, 13]. Unfortunately, many of these activities have been conducted in a vacuum, without coordination, with little or no input from stakeholders, and with little emphasis upon software deployment or on a given technique's actual effectiveness in terms of scientific understanding [9]. There have been many examples of "pretty pictures" that serve well as vehicles of communication, but such visualization results often do not necessarily produce substantive forward progress in terms of scientific understanding.

Three primary challenges face the SciDAC visualization and data understanding communities in their quest to have a substantive impact on science and engineering. First, visualization and analytics research and development must be more attuned to the needs of SciDAC scientists and

engineers. While creating scalable visualization tools is an important part of a broad strategy for meeting those needs, simply creating an image of a terabyte's worth of raw scientific data does not guarantee increased scientific insight. Indeed, if the resulting image is too complex, it may even reduce the likelihood of discovery.

Second, visualization and analytics technology must be designed and implemented as part of a larger technology ecosystem in which it is integrated with other important data management and work-flow systems to better meet the data understanding needs of the scientific community. They need to be integral with data acquisition, management, storage, and retrieval, and be a part of a researchers work-flows and systems rather as separate component parts or isolated islands of capability. Visualization and analytics technologies that are engineered to leverage complementary technologies in data management, parallel computing, and software engineering are much more likely to succeed at their intended purpose and are therefore more likely to be used by scientists as part of their day-to-day investigatory methodologies.

Third, visualization and analytics technology must be deployed, maintained, and supported in the computing environment. It is not enough simply to publish a visualization research paper and post source code on a web site. Scientific researchers need expert help from visualization and analytics scientists to deploy, tune, and adapt technology for their specific scientific domains. VACET's mission and scope directly addresses all of these important challenges.

5 Approach

Our proposed work plan directly responds to the data understanding needs of our scientific stakeholders and will produce results that meet those needs. As a Center, we are well-positioned to coordinate development and deployment activities across a spectrum of data understanding technologies and groups as well as with other SciDAC centers. This approach ensures our team's results are part of a well-balanced program within a larger technology ecosystem.

5.1 Provide Production-Quality, Petascale-Capable Visual Analysis Software Infrastructure

SciDAC's stated mission is to use advanced computing to further the DOE research programs in basic energy sciences, biological and environmental research fusion, and high-energy and nuclear physics. More specifically, the SciDAC program aims to produce "a Scientific Computing Software Infrastructure that bridges the gap between advanced research in applied mathematics and computer science, and computational science research in the physical, chemical, biological and environmental sciences." [2].

5.2 Visual Data Analysis as an Integral Part of the Scientific Process

Scientific visualization plays an important role in the scientific process: as "virtual oscilloscope" to "see the unsee-able"; as the most visible element of scientific research and insight; and as the visual component of day-to-day diagnostic and exploration tools. Its aim is to help scientists gain insight into structures, relationships, and anomalies "hidden" within data. Its impact can be seen in all areas of science, medicine, engineering, finance, security, and safety.

5.3 The Customer/Stakeholder Relationship

Our visualization team already has strong relationships with application scientists, and team members holding positions within DOE will ensure our solutions are deployed at DOE's large-scale open computing facilities, including the National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory and the Leadership Computing Facility (LCF) at Oak Ridge National Laboratory. Most members of our team have been active in conducting visualization research, development, and deployment activities during the first five years of SciDAC. We are collectively dedicated to our Center's mission: To foster scientific insight through creating and deploying effective data understanding technology that is truly responsive to the needs of the scientific research community.

The Customer/Stakeholder relationship comprises one of the foundational elements of our overall approach. These relationships, are long-term, collaborative and mutually beneficial to both stakeholder and VACET. Most of the work we perform in VACET is aimed at meeting scientific data understanding needs. Those needs are articulated by the stakeholder to use within the context of these relationships. The solutions are then tested and applied on stakeholder problems to meet scientific data understanding needs, and where appropriate, included in releases of production quality software.

5.4 Deployment Approaches

In the way there is no "one-size-fits-all" solution to scientific data understanding, there is similarly no one deployment solution that will satisfy all stakeholder operational requirements. We adopt a set of tactics, described in the following section, that affords us the framework needed to achieve maximum impact, maximum mobility and flexibility in delivering solutions, and optimizes the pipeline from concept to working software solution.

5.4.1 Applications

One of the primary strengths of our team's approach is evidenced by the fact that we are delivering production-quality solutions to customers after less than six months since VACET began business. The approach we use to accomplish this remarkable achievement is by leveraging a large body of production-quality software infrastructure as the base for both R&D and delivering new capabilities. Our two primary delivery platforms consist of VisIt and SCIRun.

The SCIRun system has been a focus of research and development at the SCI Institute since 1995 [11, 19, 21, 18, 20, 10, 16, 15, 12]. It is framework for visualization, modeling, and simulation, and has been the test bed for significant fundamental research in visualization techniques and their applications to real-world scientific problems. The strengths of SCIRun derive from its modular data flow architecture, which provides a much wider range of flexibility via modular pipelines and *dynamic compilation*. SCIRun2 [22] expands dramatically on these ideas to bring component-based scientific computing and visualization to an entirely new level. The primary novel feature in SCIRun2 is the concept of a metacomponent, [25] which allows construction of scientific software that involves mixtures of components from different sources, including support for the Common Component Architecture (CCA) [3], the Visualization Toolkit [23], CORBA [17], and dataflow components from the original SCIRun. Components from these different sources can be combined in a single computation via the use of automatically- or semi-automatically-created bridges. SCIRun2 also enables parallel components through multi-threading for shared memory programming, and parallel-to-parallel remote method invocation [4, 7] for connecting components in a distributed memory environment.

VisIt [5, 6] is a turnkey application for data exploration, code assessment, and quantitative analysis suitable for use on tera- and peta-scale datasets. In addition to standard visualization methods, the tool is used for code-to-code comparisons, code-to-experiment comparisons, analysis of parameter studies, and quantitative analysis across a variety of scientific areas. It has won an R&D 100 award, been downloaded over 25,000 times, has over 300 customers at LLNL, is used at AWE, many ASC Alliance sites, and has a large number of customers from businesses, universities, and national laboratories. Originating in 2000 as part of the ASC program, it has grown to over one million lines of code in addition to leveraging many third-party libraries. Its development has focused on key areas where solutions do not already exist: large data infrastructure, unusual data models, custom and extensible quantitative analysis and the infrastructure that binds them together. It has a scalable architecture for running expensive (I/O or compute) operations on a parallel machine to leverage resources "close to" the data and supports a client-server model for effective remote visualization use. It is extensible – developers can write plug-ins for any stage of I/O, visualization or analysis processing. Such plug-ins may run in parallel, thus providing a stable development environment for new techniques in scalable visualization and analysis.

Both tools perform standard visualization operations: contouring, volume rendering, slicing, etc., in addition to extensive subsetting, derived quantity, and analysis capabilities. VisIt has been tailored for processing large data and its server has been parallelized for distributed memory environments. As a result of this expertise and experience, our team is well positioned to respond to our science stakeholder needs.

5.4.2 Libraries and Modules/Components

We also recognize the different needs of the users in terms of delivery vehicles of new technology. Some have already committed to a particular visualization tool and need new capabilities like high performance volume rendering or feature extraction. Others require new focused tools that address specific needs such as real time monitoring of large scale simulations or remote data access and collaborative data exploration. Many users can be well served with general purpose visualization tools, such as VisIt or SCIRun, and we'll provide extensions to support new features and special data formats. We will address this diversity of needs with a three-tiered strategy with development of: **libraries** with narrowly scoped capabilities (e.g. a volume rendering library) and simple APIs that allow easy integration even if at some cost in space and/or time efficiency; **components** that can be easily combined in a data flow network exploiting the efficiency of the parts at their best; and **fully featured application visualization tools** with a large set of features that can satisfy a wide range of users.

This deployment strategy, together with extensive documentation, examples, and tutorials, will also facilitate the process of dissemination to a larger community independent of our ability to provide direct support to each of them.

5.5 Technical Portfolio

The strength of our overall technical approach is based upon a diverse portfolio of capabilities. Since there is no one panacea to the challenges of petascale data understanding, we are leveraging a set of investments across a diverse range of interrelated focus areas. These areas combine in various combinations to produce a final solution for a given stakeholder. Again, our organization as a Center affords us the opportunity to manage such a portfolio to deliver effective solutions across a wide range of science disciplines, each with slightly different needs.

The details of our technical portfolio exceed the scope of this document. In general terms, the focus areas of our technical portfolio include:

Advanced Visualization Techniques

- Integration of Fundamental Tools for Visualization and Analysis
- Publication Quality Images, Illustrations and Movies
- Project-Wide Visualization Tools
- Flow Visualization
- Scalable Visualization, Analytics and Rendering Tools
- Remote Data Access and Streaming Techniques
- Collaborative Tools

Analytics and Knowledge Discovery

- Query-Driven Visualization
- Feature Detection and Identification
- Integration of Statistical Information
- Information and Scientific Visualization Techniques
- Multifield Visualization and Analysis
- Comparative Visualization and Analytics
- Uncertainty Visualization

6 Team Organization

As described in the VACET proposal, our Center is organized into functional groups to achieve several distinct objectives: (1) facilitate the flow of information between the Center's leadership, personnel, and science stakeholders; (2) to provide the organizational structure needed to ensure oversight and coordinated operations of the Center's collection of activities; (3) to ensure we meet our work deliverables; (4) to gracefully accommodate future growth and respond to changing priorities. The Center's original functional groups are: the Center PIs, the Executive Committee, the External Advisory Board, Research and Development, Chief Software Engineer, Software Development and Support, and Stakeholder Projects. Due to a 45% reduction in budget, we have eliminated the formal External Advisory Board.

Figure 1 depicts the working interaction between these functional groups. The co-Directors (Bethel and Johnson) are responsible for the operation of the entire Center and are the point of contact between the Center and DOE. The Executive Committee (EC), which makes decisions about research, development, stakeholder applications selection and prioritization, and software deployment, is comprised of the participating Laboratory or University PIs (Bethel, Johnson, Pascucci, Ahern, and Joy).

Given the high priority on delivering useful software to our scientific stakeholders, each lab site will have a primary software engineer whose duty it is to assure software developed and deployed from that site meets the Center software engineering criteria. The set of site engineers is known collectively as the Software Engineering Group (SEG). In addition to developing, testing, documenting and maintaining the Center software, the Software Engineering Group will integrate results from all research and development groups into the Center software, relying on feedback from the project leads in the Stakeholder Projects Group, the Executive Committee and the Chief Software Engineer (Childs).

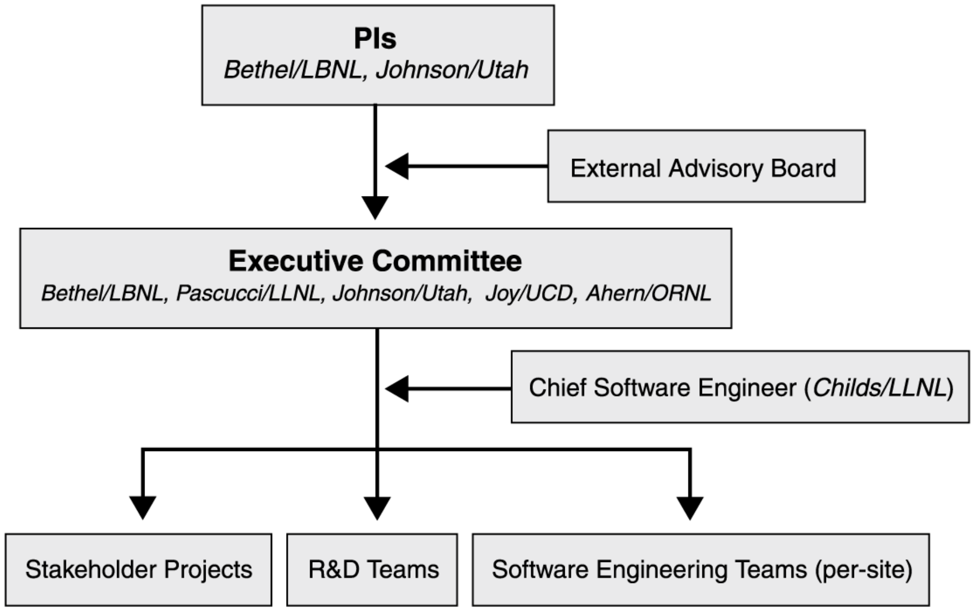

Figure 1: Functional interaction of groups within VACET.

The Chief Software Engineer (CSWE) serves many important functions within the center. One is to facilitate the coordinated design, implementation and integration of the Center's technologies into software solutions that meet stakeholder needs. He will provide guidance to the software development teams so that individual software tools and libraries will be readily usable throughout the Center's collection of stakeholder projects. He will coordinate with the EC to prioritize software development targets, and serve as a technical software advisor to the Center as a whole. He will interact with the R&D team to help foster early designs that fit well within the Center's technology implementations. He will direct the development, testing, deployment, and support of the Center software toolsets.

The Stakeholder Projects Group (SPG) is the primary interface to our science stakeholders. In this group, individuals from the Center will interact directly with science stakeholders to obtain and prioritize science needs, coordinate with the Center's EC and Chief Software Engineer to translate those needs into a work plan, to oversee and manage the work so that software is delivered to the science stakeholder.

From each participating institution is a "Site PI." That person is nominally part of the EC, and is responsible for managing issues at that organization. These issues range from responding to administrative requests, like updated budgets, as well as for managing operations of VACET staff at that organization.

Finally, the PIs and EC will work with an External Advisory Board (EAB) to obtain advice/feedback and to update them with status information. The EAB will convene once per year for a one-day meeting to review the Center's progress and strategic plans. Where feasible, the Center will bear the travel costs for the EAB members.

7 Roles and Responsibilities

7.1 LBNL

7.1.1 Bethel

As coordinating PI for VACET, Bethel's duties include overall management of a five-institution project, direct management and supervision of LBNL staff contributing to the effort, and interfacing with DOE on all matters relating to the project. In addition to project management duties, Bethel will serve as a Stakeholder Project leader, translate Stakeholder needs into project deliverables, prioritize work tasks, design and implement visualization and analytics solutions particularly in the areas of remote, distributed and parallel visualization, query-driven visualization and analysis, and latency-tolerant visualization and rendering.

7.1.2 Aragon

Cecilia Aragon is the Team Leader for the Spectrum Synthesis Visual Data analysis project for our Astrophysics stakeholders. Aragon is also Team Leader for the Nearby Supernova Factory software infrastructure, which includes data management, visualization and visual analytics. She is shared with the NERSC Analytics team, the Nearby Supernova Factory and the LBNL Visualization research effort.

7.1.3 Siegerist

Cristina Siegerist provides one-on-one consulting in the application of visual data analysis technologies to stakeholder projects. She is shared with the NERSC Analytics team.

7.1.4 Stockinger

Kurt Stockinger's background is in Scientific Data Management. As a member of LBNL's Visualization Group, and part of the VACET team, he serves as a bridge between efforts in visualization, analytics and scientific data management. He participates in several different VACET projects, and is Co-PI of a Science Application Partnership aimed at producing, applying and maintaining a high performance data model and I/O infrastructure for use by the Accelerator modeling community. He is shared with the NERSC Analytics Team and the LBNL Visualization research efforts.

7.1.5 Weber

Gunther Weber is team-leader-in-training for the Production AMR Visualization effort in VACET. He is also the Team Leader for the 3D LIC R&D effort. He is shared with the NERSC Analytics Team and the LBNL Visualization research efforts.

7.2 ORNL

7.2.1 Ahern

Sean Ahern is the VACET PI for Oak Ridge National Laboratory (ORNL). In addition to directly managing the contributions of the personnel at Oak Ridge to VACET, Ahern will serve as a direct customer liaison to users sited at ORNL, as well as users of ORNL's Leadership Computing Facility (LCF). In this role, Ahern will translate Stakeholder needs into VACET deliverables, prioritize ORNL work tasks, design and implement visualization and analytics solutions particularly where

they impact users of the LCF. Specific ORNL customers are likely to be Dr. Don Batchelor in fusion simulation and Dr. John Drake in climate simulation.

7.2.2 Ostrouchov

Dr. George Ostrouchov is a statistician working for the Computer Science and Mathematics division of ORNL. In his role in VACET, Dr. Ostrouchov will serve as a liaison to customers in the area of statistical analysis, feature detection, and coupled statistics/visualization deployment. He will directly liaise with users located at ORNL, specifically Dr. Don Batchelor's fusion project. Dr. Ostrouchov also has continuing collaborations with the Scientific Data Management Center for Enabling Technology, allowing for cross-fertilization between the two centers.

7.2.3 Meredith

Jeremy Meredith is a computer scientist working for the Computer Science and Mathematics division of ORNL. He has extensive experience working with distributed parallel visualization and data analysis codes, especially the VisIt visualization system. He will serve as one of the primary code developers for the VisIt system, integrating new analysis capabilities, and customizing data readers for various VACET customers.

7.3 LLNL

7.3.1 Pascucci

Valerio Pascucci is the coordinator for VACET activities at Lawrence Livermore National Laboratory (LLNL) and will have responsibility for direct management and supervision of LLNL scientists involved in the project. In this capacity, Valerio will translate and prioritize application stakeholder needs into LLNL project deliverables, design and implement data analysis and visualization solutions particularly in the areas of robust topological analysis, feature extraction and tracking, linked views, remote, and distributed streaming techniques.

Valerio will also serve as a Technical Point of Contact for the climate modeling application area and will work with Dean Williams to identify VACET deliverables to be enacted in collaboration with the Earth Science Grid and with John Drake to identify VACET technology that can directly support climate modeling science applications.

7.3.2 Childs

Hank Childs is the VACET Chief Software Architect and is responsible for the implementation of the activities defined by the Executive Committee with respect to software deployment and distribution as well as quality control.

Hank Childs is also the main LLNL developer responsible for introducing new features in VisIt, such as full support AMR data to satisfy the needs of the APDEC center and relative science applications.

7.3.3 Bremer

Peer-Timo Bremer is a Computer Scientist at LLNL. He will be the main person responsible for developing robust topological techniques for data analysis and feature extraction for structured and unstructured data. Peer-Timo will work with scientists in different applications to develop formal

topological characterization of features of interest and corresponding implementation of discrete algorithms for their computation while providing guaranteed error bounds.

7.3.4 Laney

Daniel Laney is a Computer Scientist at LLNL. He will be the main person responsible for management of large scientific datasets with particular focus on multiresolution streaming techniques. In particular, Daniel will work on the use of progressive techniques for the analysis and visualization of large models from petascale science applications.

7.3.5 Mascarenhas

Ajith Mascarenhas is a Postdoc at LLNL. He will be the main person responsible for developing robust topological techniques for time dependent datasets and will work on reliable feature tracking methods. His work will include advanced techniques enabling the use of a multi-scale notion of time and allow performing direct selection and filtering of features of in a space-time continuum.

7.3.6 Bonnell

Kathleen Bonnell is a Computer Scientist at LLNL who has extensive expertise developing scientific visualization techniques and who has a deep understanding of the VisIt visualization platform. Kathleen will be responsible for integrating a number of new visualization components and data readers within VisIt as well separating particular VisIt portions into stand-alone libraries that can be deployed separately in other environments.

7.4 Utah

7.4.1 Johnson

Professor Chris Johnson is the VACET Co-PI. Along with VACET Co-PI Wes Bethel, he will lead the overall VACET effort. In addition, Johnson will help create new visualization techniques for displaying error and uncertainty, as well as new multi-field visualization techniques.

7.4.2 Hansen

Professor Hansen and a graduate student will be working on multi-field visualization and uncertainty visualization, focusing on GPU methods for volume rendering for multi-fields taking uncertainty into consideration.

7.4.3 Silva

Professor Claudio Silva will pursue the deployment of VisTrails in the climate application for the purpose of comparative analysis. He will be working with Valerio Pascucci at LLNL, who is the direct contact with Dean Williams (climate stakeholder).

7.4.4 Parker

Professor Steven Parker will work on adapting Utah's real-time ray tracer to particle-based datasets (stakeholders in fusion, astrophysics, accelerator modeling). He will also be working with the CPES center to implement a web-based visualization tool to automate routine visualization tasks for the

two-dimensional XGC simulation code. Using Web 2.0 technologies, we will enable viewing large numbers of particles and mesh points in a normal web browser.

7.4.5 Sanderson

Dr. Allen Sanderson's role in VACET will be two-fold: (1) a liaison between VACET and fusion and astrophysics scientists addressing their needs as a stakeholder; (2) developing and deploying visualization and analysis tools based primarily on the needs of fusion and astrophysics stakeholders. During the first year, Sanderson will focus primarily on deploying GPU and volume rendering techniques for visualizing and analyzing particle-based datasets.

7.4.6 Tricoche

During the first year, Dr. Xavier Tricoche will focus on multi-field and flow visualization algorithms that target the visual understanding needs of multiple stakeholders. He will contribute software implementations of flow visualization topology-based algorithms in SCIRun and VisIt. He will also work towards the characterization and the representation of salient structures in high-dimensional spaces resulting from the embedding of multiple related quantities in a single analysis space.

7.4.7 Cole

Martin Cole is a staff software engineer who will work on transitioning visualization research algorithms to end user software.

7.5 UC Davis

7.5.1 Joy

Joy's duties involve the interaction with stakeholders to identify fundamental visualization research and development problems, to design and translate these problems into workable solutions, and to develop deliverables that can be returned to our stakeholders to address their data analysis needs. His initial efforts will focus on highly-parallel methods for material interface reconstruction, query-driven visualization, and multi-dimensional visualization. These efforts are targeted at stakeholders in combustion, astrophysics, and applied mathematics.

7.5.2 Hamann

Hamann's duties involve investigating topological solutions and multiresolution methods to support analysis and exploration of petascale data sets. He will focus on out-of-core methods and visual segmentation methods, which allows one to represent massive data sets by a relatively smaller number of elements, and implementing these techniques within the frameworks supported by VACET. Through these efforts, he will collaborate with VACET's LLNL component (Pascucci), to adapt and deploy these methods to our VACET stakeholders.

8 Interrelationships with Other Projects – Centers, Institutes, Applications

8.1 Science Applications

8.1.1 Accelerator

- R. Ryne, LBNL. *Advanced Computing for 21st Century Accelerator Science and Technology.* VACET team members contribute to a Science Application Partnership project that focuses on a community-wide data I/O library. The overall objective is to foster a community data standard so that researchers can share data between teams, as well as between different codes in an ensemble model. An additional benefit is reduced effort for applying advanced visual data understanding technology to this science area. This effort is expected to continue in a new Science Application Partnership with the Community proposal, currently under review at DOE for a future award.
- P. Spentzouris, FNAL. *Proposed Community Petascale Project for Accelerator Science and Simulation (COMPASS)* This prospective project, which has not yet been selected for an award, represents a major future stakeholder in Accelerator science technology. We have engaged in several early stakeholder needs assessment meetings with several principals from this team. Many of their needs overlap with other those of other stakeholders, notably, those projects in Fusion that generate particle-based data.

8.1.2 Astrophysics

- S. Woosley, UC Santa Cruz. *Computational Astrophysics Consortium: Supernovae, Gamma Ray Bursts, and Nucleosynthesis.* VACET is providing production quality, AMR visual data analysis infrastructure (Section 12.6.1) as well as performing R&D for new visual analytics techniques appropriate for Spectrum Synthesis analysis (Section 12.2.1).

8.1.3 Climate

- J. Drake, ORNL. *A Scalable and Extensible Earth System Model for Climate Change Science.* Drake is the consumer of new technologies for climate data analysis and visualization.
- D. Randall, Colorado State. *Design and Testing of a Global Cloud-Resolving Model.* Randall and his team are a prospective VACET collaborator. A member of Randall's team attended the VACET AHM in Salt Lake City in February, 2007.

8.1.4 Combustion

- John Bell and Marc Day, LBNL/CCSE. Bell, a member of the National Academies of Science, performs basic combustion research under DOE base program funding in Mathematics. He is a customer of Colella's Applied Partial Differential Equations Center. Our work with Bell focuses on developing and deploying production quality AMR visual data analysis infrastructure (Section 12.6.1) as well as advanced visual analysis capabilities aimed at gaining traction on difficult scientific data understanding problems (Sections 12.4.1 and 12.4.3).
- Jacqueline Chen and Evatt Hawkes, SNL-CA. *Direct Numerical Simulation of Turbulent Non-Premixed Combustion – Fundamental Insight Towards Predictive Modeling.* VACET's interactions focus on R&D of techniques for topologically based feature identification, tracking and visualization. Traditional visualization methods to not provide the kind of hard, quantitative information needed to advance combustion science. Chen's work is conducted under DOE's INCITE program.

8.1.5 Fusion

- S. Ethier and W. Lee, PPPL. *Center for Gyrokinetic Particle Simulations of Turbulent Transport in Burning Plasmas.* In this active project, VACET is providing novel visual data analysis tools for examining orbitals in very large particle-based datasets. In new work, which is the subject of a May 2007 SAP proposal, we would focus on applying query-driven visualization techniques to analysis of large MHD datasets, and including visual data analysis techniques as part of scientific workflow (this work would be joint with the SDM Center).
- S. Kruger, Tech-X. *Center for Extended Magnetohydrodynamic Modeling.* VACET is providing novel visual data analysis tools for use on accelerating scientific understanding of data produced by MHD codes.
- D. Batchelor, ORNL. *Simulation of Wave Interactions with Magnetohydrodynamics.* This active project has made progress in topological analysis of magnetic field lines and a data conversion utility for loading simulation data into VisIt.
- J. Cary, Tech-X. *Framework Application for Core-Edge Transport Simulations.* This potential stakeholder project is still early in its design stage and is not quite ready as of the time of this writing for a collaboration with VACET. A representative from VACET attended the FACETS project kickoff meeting in Boulder, CO in December 2006.
- P. Bonoli, MIT. *Numerical Computation of Wave-Plasma Interactions in Multi-dimensional Systems.* This prospective VACET project, which is the subject of May 2007 SAP proposal, would focus on providing fundamental visual data analysis infrastructure and expert assistance applying it to fusion data analysis, visual data analysis of phenomena in the confinement chamber: particles, field lines and scalars.
- S. Jardin, PPPL. *Center for Extended Magnetohydrodynamic Modeling.* Presently, VACET team member A. Sanderson (Utah) is on an existing SAP with CEMM. New work, which is the subject of a May 2007 SAP proposal, we would continue work on topological analysis of Poincaré plots to automatically find islands, and provide fundamental visual data analysis infrastructure to replace aging (and no longer supported) tools presently in use at PPPL for day-to-day analysis.
- W. Nevins, LLNL. *The Plasma Microturbulence Project.* This prospective project, which is the subject of a May 2007 SAP proposal, would focus on tools and techniques for visual data analysis of particle-based data at multiple scales and field line integration.

8.2 Centers

- Dean Williams, LLNL. *Scaling the Earth System Grid to Petascale Data Center for Enabling Technologies.* Dean's team produces software infrastructure that is widely used in the climate research community for data storage, haring, analysis and visualization. VACET's new technical contributions will be deployed through CDAT to achieve broadest possible market penetration.
- P. Colella, LBNL. *The Applied Partial Differential Equations Center for Enabling Technologies.* Colella and his team are one of VACET's primary stakeholders. VACET is working to provide a production-quality replacement for the older ChomboVis application. The impact to APDEC is a significant cost savings and the ability to perform visual analysis of large and complex AMR data on parallel machines.
- A. Shoshani, LBNL. *The Scientific Data Management Center.* We have identified several stakeholder projects where SDM's FastBit index/query technology would be helpful for accelerating visual analytics applications, and where VACET's visual analytics capabilities would be included in scientific workflows. These are "future work" items.

- D. Skinner, LBNL. *The SciDAC Outreach Center.* VACET has provided input to Skinner concerning VACET needs of the Outreach Center. We have highlighted the need for support for collaborative software and release engineering capability, as well as collaborative documents, spreadsheets and project management tools. As a result, Skinner is including GForge[2] as part of the Outreach Center's infrastructure offerings.

8.3 Institutes

- J. Owens and K.-L. Ma, UCD, for the *SciDAC Institute for Ultrascale Visualization.* Active, Joint project studying use of collections of GPUs to accelerate large data rendering.
- W. Kramer, LBNL, for the Petascale Data Storage Institute. Prospective VACET interactions would focus on joining forces with PDSI on data I/O infrastructure, models and formats for petascale applications.

8.4 Partnerships

- J. Bell, LBNL. Bell is PI of the Science Application Partnership entitled *Computational Astrophysics Consortium: Adaptive Algorithms*, which is part of Woosley's Astrophysics Consortium. VACET interactions with Bell span two different focus areas. First, VACET is working to provide to production-quality AMR visual data analysis software infrastructure to Bell's team (Section 12.6.1). Second, VACET is engaged in focused research projects that address specific scientific data understanding challenges.
- K. Stockinger, LBNL. Stockinger is Co-PI of a Science Application Partnership entitled *H5part: High Performance Data I/O for the Accelerator Community*, which is part of Ryne's concluding SciDAC Accelerator project (Section 8.1.1). By interacting with this project, which aims to develop and deploy an Accelerator community-wide data model and I/O capability that includes integration of index/search technology from the SDM Center, we are well positioned to quickly achieve dramatic results in the Accelerator community when the new Accelertor SciDAC projects begin in mid-2007. A follow-on Partnership entitled *High Performance Parallel Data Management and Analysis Tools for Beam Dynamics Modeling* is part of the P. Spentzouris Accelerator proposal currently under consideration for a SciDAC award.

8.5 Facilities

- NERSC and its Analytics Team[3], which provides front-line visual data analysis and analytics support to projects hosted at NERSC. The scope of that Team's effort includes data management, workflow management, visualization, analysis and analytics. VACET works closely with members of NERSC Analytics in several different capacities: we share common stakeholders; NERSC Analytics is a consumer of VACET technologies and assistances; through ERCAP, NERSC provides access to computational and storage resources for use by VACET in fulfilling its production visual data understanding mission.
- LCF/ORNL blurb goes here (Sean, this is your baby)

8.6 Other Research Efforts

- LBNL Visualization Group[4], which has ongoing R&D projects in the areas of high performance Query-Driven Visualization and Analytics.

[2] http://www.gforge.org
[3] http://www.nersc.gov/nusers/analytics
[4] http://vis.lbl.gov/

- Nearby Supernova Factory[5]. The SNfactory is an experiemnt to develop Type Ia supernovae as tools to measure the expansion history of the universe and explore the nature of Dark Energy. It is an international collaboration between several groups in the US and France. It is a "proving ground" for the younger Supernova Acceleration Probe (SNAP) experiment[6]. The SNfactory team will directly benefit from our work on Spectrum Synthesis analysis. One of our stakeholder projects in this subject area, P. Nugent (LBNL) for S. Woosley, is part of the SNfactory team. In addition, a VACET team member, C. Aragon (LBNL) is the team leader for the SNfactory data management and analysis infrastructure effort.

8.7 INCITE

- Warren Mori, UCLA. *Petascale Particle-in-Cell Simulations of Plasma Based Accelerators.* Prospective VACET stakeholder, joint with the NERSC Analytics team.
- Chuang Ren, Rochester. *Three-Dimensional Particle-in-Cell Simulations for Fast Ignition*, Prospective VACET stakeholder, joint with the NERSC Analytics team.
- Jacqueline Chen and Evatt Hawkes, SNL-CA. *Direct Numerical Simulation of Turbulent Non-Premixed Combustion – Fundamental Insight Towards Predictive Modeling.* VACET's work with Chen is described above in Section 8.1.4.

9 Change Control

Over the course of any project, changes are inevitable. Our project's mission statement – "to foster scientific insight through creating and deploying effective data understanding technology that is truly responsive to the needs of our stakeholders in the scientific research community" – will likely require constant tuning, adaptation and evolution of VACET's approach, scope, operational model. Change is inevitable in light of the fact VACET's customers will change, the technology environment will change, new opportunities will arise, DOE programmatic priorities will change, and so forth. Therefore, our overall objective here is to optimize the benefits created by VACET while minimizing the adverse impact of change. To a large extent, success in this regard will require careful management of project scope and focusing resources in areas with the greatest impact.

In this section, we discuss briefly our approach to Change Control for several different scenarios in which project change can be reasonably anticipated.

9.1 Engaging with Customers

The decision to engage or not with a particular new stakeholder project is a "business decision" that requires the approval of the project PIs. The mechanism we use for evaluating such a decision is as follows: (1) the existence of an opportunity for work with a new stakeholder is brought to the attention of the Executive Committee; (2) the potential cost/benefit of working with that stakeholder is discussed amongst the EC via email and/or at one of the regularly scheduled EC calls; (3) the EC makes a recommendation; (4) the PIs render a decision taking into account the EC's recommendation.

In some instances, there are factors external to VACET that would influence the decision to engage or not with a particular stakeholder. Such factors would include input from outside sources, like ASCR/DOE.

[5] http://snfactory.lbl.gov
[6] http://snap.lbl.gov

9.2 Scope of Customer Project

A "typical" customer/stakeholder project involves first ascertaining stakeholder needs, translating those needs into tasks or deliverables, selecting some subset of those deliverables for VACET action, performing the work to produce the deliverable, and iterate. Other aspects of this lifecycle, such as long-term maintenance and support of the deliverable, are omitted from the discussion here for brevity.

Once a decision is made to proceed with a specific customer, the Customer Project Manager is responsible for generating the list of stakeholder needs and translating those needs into tasks or deliverables. The decision of whether or not to proceed with any/all of those tasks is also a "business decision" requiring EC review and recommendation and PI approval. This decision is important because it commits VACET resources, and often requires coordinating efforts across a number of different teams. Changes to scope – e.g.,., adding or removing deliverables – must be subject to EC review and recommendation and PI approval.

9.3 Project Management Plan

This Project Management Plan serves several purposes. It reflects the VACET mission, approach, benefits, etc. – it is a "business plan." As it is regularly updated, it is a record of accomplishments. It also presents a snapshot of goals and objectives for the future. As such, changes to "the business plan" as well as the list of future work targets requires EC review and recommendation and PI approval. The list of accomplishments are provided by all team members and reviewed by the EC. The PIs approval on the PMP is implicit due to the fact the PIs are responsible for generating the PMP with EC oversight and input from the entire team.

10 Center-Wide Accomplishments

10.1 Stakeholder Projects

In its first six months of operation, VACET has begun meaningful, high-impact work with stakeholder projects in a number of different science areas.

10.1.1 Accelerator Modeling

One of our stakeholder projects aims at applying high-quality interactive rendering techniques applied to particle-based datasets, such as those from PIC-based codes to perform Plasma-Wakefield accelerator modeling (Section 12.1.1). The primary stakeholder for this project is the VORPAL development team at Tech-X Corporation[7].

10.1.2 Astrophysics

As documented on the VACET twiki[8], our primary Astrophysics stakeholder is Stan Woosley's Computational Astrophysics Consortium SciDAC Science Application[9]. Our interactions with that team focus on two primary projects.

The first is providing production quality AMR visual data analysis infrastructure in support of John Bell, who is responsible for creating the AMR-based computational infrastructure for modeling

[7]http://www.txcorp.com/products/VORPAL/
[8]http://sci.utah.edu/vacetwiki
[9]http://www.scidac.gov/physics/grb.html

Supernova explosions. VACET is achieving some economy of scale in this project by grouping together John Bell's needs with those of Phil Collela's APDEC Center, which also focuses on R&D of AMR-based computational infrastructure. The needs of these respective projects overlap to a large degree, so our overall objective is to provide production-quality AMR visual analysis infrastructure to both teams. The specific needs of these groups are documented on the VACET twiki[10]. Specific progress on the VACET AMR production-quality visual data analysis infrastructure is documented below in Section 12.6.

The second is in developing tools and techniques for replacing "Chi by eye" supernova spectrum/light curve comparative analysis. This effort has special challenges and needs, which are documented on the VACET twiki[11]. Work in this area will be applicable to future stakeholder projects where some form of machine learning/classification methods will be useful to discover relationships in large, multidimensional datasets. This particular project is groundbreaking in several respects. First, replacing "Chi by eye" with rigorous, quantitative methods will represent a major leap forward in data understanding capabilities. Second, the intersection of visual analytics with analysis and visualization on such problems represents a major growth area in our field. Third, this type of approach is promising for providing traction on Petascale and beyond data understanding in a way not possible when using only visualization or only data analysis techniques. This project's accomplishments, plans and risks are listed in Section 12.2.1.

10.1.3 Climate

VACET's multi-faceted approach to addressing the needs of the climate community is as follows. First, we are working closely with Dean Williams, PI of the Earth Systems Grid SciDAC project[12]; Dean is the lead on the Climate Data Analysis Toolkit (CDAT)[13], which is widely used in the climate community for data analysis and visualization. VACET's approach here is to achieve maximum possible market penetration into the climate research community by providing new capabilities in one of its most widely used software toolkits for data management, analysis and visualization. The second approach is to form close collaborations with specific stakeholders. At the present time, our climate stakeholder is John Drake, ORNL, who is PI on the new SciDAC project "A Scalable and Extensible Earth System Model for Climate Change Science"[14]. Both Williams and Drake have provided to VACET a set of "stakeholder needs" that form the basis for current work activities. These needs are documented on the VACET twiki[15].

Two active projects focus on (1) improving CDAT user productivity while enabling new forms of comparative visualization and analysis (Section 12.3.1); (2) deploying new advanced visualization capabilities in CDAT for use by the climate community (Section 12.3.2).

10.1.4 Combustion

VACET has launched a number of interrelated projects for combustion stakeholders. One major thrust is R&D and deployment of production quality AMR visual data analysis software infrastructure (Section 12.6.1). Others are aimed at R&D of new capabilities needed by combustion researchers to understand their data. These projects include topologically based feature identifi-

[10]http://www.sci.utah.edu/vacetwiki/index.php/Collab:AMR_Applications
[11]http://www.sci.utah.edu/vacetwiki/index.php/Collab:Astrophysics
[12]http://www.scidac.gov/compsci/ESG.html
[13]http://www-pcmdi.llnl.gov/software-portal/cdat
[14]http://www.scidac.gov/climate/earth.html
[15]http://www.sci.utah.edu/vacetwiki/index.php/Collab:Climate_Modeling

cation, tracking and analysis (Sections 12.4.1 and 12.4.2); and multivariate, spatiotemporal visual data analysis 12.4.3); advanced techniques for vector field visualization (Sections 12.9.1, 12.9.2).

10.1.5 Fusion

VACET has a number of focused R&D projects, each of which is tackling a different aspect of a larger set of data understanding needs that arise in Fusion. These new techniques are being designed, implemented and tested in one or both of our production delivery platforms, SCIRun and VisIt.

One project focuses on visual analysis of particle orbits in very large particle-based datasets (Section 12.5.1). Another performs topological analysis of Poincaré plots to identify formation of islands in magnetic fields (Section 12.5.2) in MHD simulation output.

Yet another project explores use of linked views and a parallel coordinates display and selection mechanism to provide the ability to rapidly select subsets of large and complex data for analysis on the basis of combinations of data ranges across a number of variables (Section 12.8.1). While initially motivated by the needs of the multiple Fusion Microturbulence projects (i.e., particle-based datasets), the concepts are generally applicable to any type of large and complex dataset. It is especially noteworthy that the results of this latter R&D project were implemented in VisIt in late 2006/early 2007, integrated into the VisIt code base and made part of a public VisIt release in April 2007.

A challenge faced by many particle-based codes (GTC, Vorpal, etc.) is the fact that a complete data dump may not be possible nor practical. Current approaches dump every n'th particle, which may or may not be a representative sample. One of our projects focuses on developing suitable statistical sampling techniques by leveraging the "R" package to help provide a better mechanism for downsampling during I/O, one that is based on a statistical model rather than one that is completely arbitrary (Section 12.8.2).

10.1.6 Mathematics

VACET's primary stakeholder in this space is the P. Colella's APDEC center. Our team is aiming to provide APDEC with a wholesale replacement for the older ChomboVis application. Our replacement, VisIt, is production-quality visual data analysis infrastructure that is suitable for use on parallel machines and on AMR data. Over the course of the next several years, we have a staged sequence of R&D plans that will result in software releases containing a lengthy set of features/capabilities needed by APDEC, as well as those APDEC support. By fostering this close relationship with APDEC, VACET will be able to achieve impact across a broader range of science applications than it would by itself.

VACET projects include: (1) a large software research, development and engineering effort that involves many of the VACET family (Section 12.6.1); (2) a subproject focusing on R&D of techniques for extraction of Embedded Boundaries in multifluid or multiphase problems on block-structured adaptive grids (Section 12.6.3); (3) a concerted effort to enable launching a visual data analysis tool from the debugger while running an AMR application (Section 12.6.2); (4) a team-wide software engineering effort to provide a library form of a state-of-the-art volume rendering engine, and integrate it into our delivery platforms (Section 12.7.1)

10.2 Forward-Looking Projects

After interactions with many stakeholders, we quickly realized that there are many needs we can anticipate in the future based upon current trends in computer and computational science. To be

well prepared to meet what we consider to be imminent future needs, we have undertaken a number of focused R&D projects aimed at pathfinding in a number of areas we feel will be crucial in the near term.

One project aims to explore techniques for discovery and visual display of interrelationships between variables in a large and complex dataset (Section 12.8.3).

Another is to develop scalable statistical techniques suitable for use on parallel platforms, for use in visualization as well as analysis Comparative Visualization/Analytics/Analysis (Section 12.8.4).

A trend of increasing importance is the notion of comparative visualization and analysis. Our effort here aims to move well beyond "Chi-by-eye" and A vs. B comparisons by moving into ensemble-level comparison, as well as comparison of data from different types of grids (Section 12.8.5).

One promising approach to visual analysis of highly multivariate datasets is the so-called "Function Field", where an $n-$dimensional space is mapped to a one-dimensional field (Section 12.8.6).

VACET researchers, in collaboration with researchers from the Institute for Ultrascale Visualization are working to develop a new system which enables the rendering of very complex, out-of-core scenes with unbiased global illumination. The system facilitates the rendering of hundreds of millions of triangles, gigabytes of texture, and complex shading by forcing high levels of coherency. (Section 12.9.4).

Related, our team is pursuing algorithmic R&D and software engineering for highly scalable platforms. One project focuses on techniques for implicit surface reconstruction in a way amenable to parallelization on CPU or GPU platforms. Another aims to provide the ability to perform multivariate volume visualization on GPU-based platforms using GLSL (Section 12.9.3).

10.3 Center-Wide Projects

10.3.1 Particle Data Focus Group

As described above, we have several active stakeholder projects that address the general theme of visual data analysis of particle-based datasets: Accelerator Modeling (Section 12.1.1), Fusion (Sections 12.5.1, 12.8.1). We expect to begin operations with at least two new stakeholders in the months ahead (Mori, UCLA and Ren, Rochester; both are INCITE awardees) who use a PIC code (Vorpal) to model Plasma-Wakefield accelerators; this PIC code generates large particle-based datasets.

While each of these customer/stakeholder projects has a corresponding VACET Customer Project Manager, all efforts are "coordinated" through a central VACET Particle Data Focus Group. Here, all persons working on visual analysis of particle-based datasets exchange information and ideas through conference calls and email. We look for opportunities to share ideas and technologies across these different projects.

A good example of such a shared technology, which has been implemented and deployed in one of our production delivery platforms, is the work with parallel coordinates to perform subset selection (Section 12.8.1).

10.4 Software Engineering Team

As described earlier in Section 6, VACET has a Software Engineering team comprised of representatives from all participating institutions. This team is tasked with centralized coordination of creating well-engineered software in the form of applications, libraries and modules. The team has held regular meetings and calls, and has been active in producing new software tools needed by our stakeholders.

The team has made excellent progress on a large, Center-wide activity aimed at providing a high-quality volume rendering capability for use in multiple deployment vehicles (e.g., Section 12.7.1). This new capability will have an impact in three different delivery platforms for at least six different stakeholder projects spanning diverse scientific disciplines.

10.5 Other Collaborations

In addition to the productive relationships with Science Application stakeholders, VACET has active and gestating collaborations with other SciDAC Centers, Partnerships and Institutes.

- **SDM Center.** VACET coordinates activities with the SDM Center to look for opportunities to share technology. VACET is presently evaluating several projects where the SDM Center's FastBit index/query technology could be used to accelerate visual analytics operations on large data. VACET is also coordinating with the SDM Center to join forces on at least two Fusion Science Application Partnership project proposals that will be submitted in May 2007. Areas of overlap in those projects include index/query and visual analytics as part of scientific workflows.
- **Visualization Institute.** We have had multiple positive interactions with IUSV. IUSV hosted a workshop at SC06 and invited several speakers from VACET. We also have engaged in a joint project involving use of many GPUs to address large-scale rendering problems (Section 12.9.4).
- **Petascale Data Storage Institute.** One of the focus areas of the PDSI is to improve high performance I/O for Petascale applications. Through PDSI participants at NERSC, VACET has engaged in preliminary discussions centering around the theme of working with PDSI on data I/O improvements for petascale applications. VACET's interest in this space is on data models and formats, since our infrastructure will be used for visual data analysis of data produced by the petascale applications. By becoming involved early in formulating and executing an I/O solution for those applications, VACET will be in a good position to rapidly achieve results as those applications come online and begin producing data.

10.6 Infrastructure

During the first six months of operations, we have begun the process of securing access to computational and storage infrastructure for use by VACET team members in fulfilling their mission. Our primary focus thus far has been on securing access to NERSC/LBNL and LCF/ORNL for the purposes of: (1) ensuring that our technologies are operational on these platforms, since they are one of our primary deployment targets; (2) beginning to deploy our technologies at these facilities for use on stakeholder data, which tends to be primarily located at one of these two facilities.

Through the ERCAP process, we have been awarded 20,000 hours and 50 TB of storage at NERSC for AY2007. Approximately a dozen VACET members now have accounts at NERSC (all VACET members are eligible for an account at NERSC).

At LCF/ORNL, S. Ahern has provided accounts on the ORNL visualization cluster to VACET team members. He has also submitted a proposal to gain access to the primary compute facilities at LCF.

Joy at UCD has purchased a small amount of equipment needed to support VACET R&D activities.

10.7 Communication

A summary of our "communication" accomplishments is as follows:

- The VACET EC has engaged in regular phone conferences and email discussions to ensure coordinated management and execution of all VACET activities.
- The VACET Software Engineering team has engaged in regular calls and email exchanges to coordinate and execute Center-wide software projects.
- The VACET Particle focus group has been active in coordinating activities related to visual data analysis of particle-based datasets.
- VACET team members have participated in dozens of face-to-face meetings with active and prospective stakeholders and collaborators.
- The VACET website, www.vacet.org, was brought online and has been undergoing regular updates.
- The VACET wiki[16] has been brought online and is an indispensable part of our Center's operations.
- Two All-Hands meetings. The first was in October 2006 in conjunction with IEEE Visualization 2006. The second was in late January 2007 at Salt Lake City, UT.
- HPCwire interview[17] with VACET personnel.

10.8 Publications

- V. Pascucci, G. Scorzelli, P.-T. Bremer, and A. Mascarenhas, "Robust On-line Computation of Reeb Graphs: Simplicity and Speed" ACM Transactions on Graphics (SIGGRAPH 2007), to appear.
- G. Weber, S.E. Dillard, H. Carr, V. Pascucci, and B. Hamann, "Topology-Controlled Volume Rendering", IEEE Transactions on Visualization and Computer Graphics Vol. 13, No. 2, pp. 330-341, 2007.
- J.C. Anderson, L.J. Gosink, M.A. Duchaineau, and K.I. Joy, "Feature Identification and Extraction in Function Fields," Proceedings of EuroVis 2007, to appear.
- J.C. Anderson, L.J. Gosink, M.A. Duchaineau, and K.I. Joy, "Exploration of Function Fields using Multiple Probes," IEEE Visualization 2007, submitted.
- L.J. Gosink, J.C. Anderson, E.W. Bethel, and K.I. Joy, "Variable Interactions in Query Driven Visualization," IEEE Visualization 2007, submitted.
- L.J. Gosink, J.C. Anderson, and K.I. Joy, "Variable Interactions in Query Driven Visualization," IEEE Visualization 2007, submitted.
- H. Childs, S. Ahern, J. Meredith, M. Miller, and K.I. Joy, "Comparative Visualization Using Cross-Mesh Field Evaluations and Derived Quantities," *IEEE Transactions on Visualization and Computer Graphics*, submitted.
- M. Miller, H. Childs, J. Clyne, G. Ostrouchov, S. Ahern, and K.I. Joy, "Equivalence Class Functions: A New and Versatile Framework for Visualization at the Extreme Scale," IEEE Visualization 2007, submitted.
- *This list is not complete. VACET team members collectively made approximately 10-12 submissions to IEEE Visualization 2007. We were unable to collect data about all of these submissions in time for inclusion in this PMP.*

[16]http://www.sci.utah.edu/vacetwiki/index.php/Main_Page
[17]http://www.lbl.gov/CS/Archive/news110306.html

10.9 Outreach, Presentations, Education and Service

- Multiple VACET presentations at the SC06 Workshop on Ultrascale Visualization: *Query-Driven Visualization Accelerates Scientific Insight* by E. W. Bethel; *VisIt Overview* by H. Childs; *Robust Topology-Based Analysis of Large Scale Data* by V. Pascucci.
- VACET team members serve in a leadership role for IEEE Visualization 2007. K. Joy is the General Chair. C. Hansen, E. Wes Bethel and V. Pascucci are on the Program Committee.
- SciDAC 2007 Program Meeting. E. Wes Bethel is on the Organizing Committee for this year's SciDAC 2007 Program meeting.
- ASCR Visualization and Analytics Workshop 2007. Chris Johnson and E. Wes Bethel are among the co-organizers of the ASCR Visualization and Analytics Base Program Workshop 2007.
- X. Tricoche delivered a presentation entitled *Characterizing the Topology of a Hamiltonian System Exhibiting Chaos: From the Standard Map to the Tokomak* at the University of Kaiserslautern.
- SciDAC Review Journal – Spring 2007. Several images appear in the Spring 2007 issue of the SciDAC Review Journal, including the image on the journal cover (see Figure 2. These images were produced by C. Siegerist, LBNL, in a collaborative effort between VACET and the NERSC Analytics team in support of science stakeholders with INCITE awards hosted at NERSC.
- *Visual Computing: Research Challenges, Harvard University, Boston, May, 2007.*
- *Large-Scale Bioimaging and Visualization*, IEEE International Parallel and Distributed Processing Symposium (IPDPS), Long Beach, March 2007 (Keynote Speaker).
- *Visualizing the Future*, Utah Computer Society, Salt Lake City, February 2007 (25th-Anniversary Keynote Presentation).
- *Computational Bioimaging and Visualization: Challenges and Opportunities*, Center for Computational Molecular Biology, Brown University (Distinguished Lecture), December 2006.
- *Visualizing the Future*, Second International Symposium on Visual Computing (ISVC06), Lake Tahoe, November 2006 (Conference Banquet Speaker).
- *Visual Computing: Research Challenges*, Harvard University, Boston, May, 2007.
- *Inverse Bioelectric Field Problems*, Inverse Days 2006, Tampere, Finland, December 2006.
- *Visualizing Uncertainty*, Uncertainty Workshop, Society of Exploration Geophysicists Annual Meeting 2006, New Orleans, October 2006.
- *This list is not complete.*

11 Center-Wide Plans

11.1 Stakeholder Projects

Each Customer Project Manager has engaged in a substantial amount of project planning activity. The details of each project's planned work activities are listed below under Section 12.

11.2 Other Collaborations

- VACET team members are contributing to five different Science Application Partnership proposals, four to Fusion and one to Accelerator. The scope of new work activities will be refined depending upon which of these are awarded and to what level.

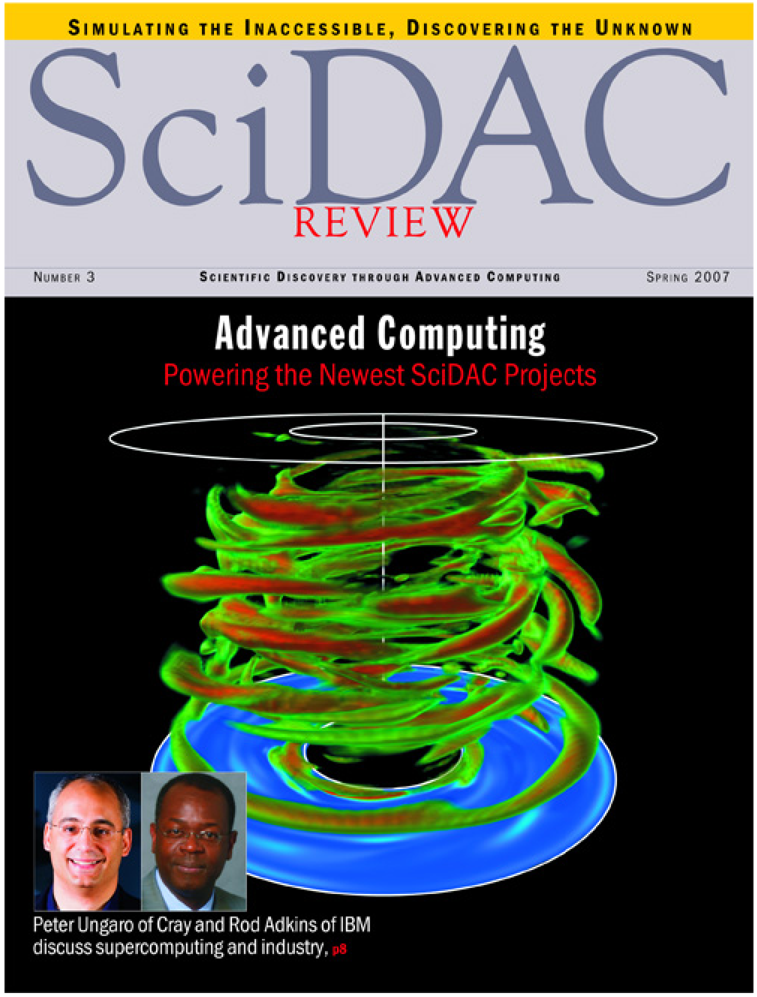

Figure 2: The cover of the Spring 2007 issue of the SciDAC Review highlights an image created in a collaborative effort between VACET and the NERSC Analytics team. It shows one time step of the evolution of total advective radial flux of axial moment, a study important in aiding understanding of the science of star and black hole formation.

11.3 Infrastructure

- VACET will submit a renewal ERCAP request for continued access to NERSC facilities in AY2008.
- ORNL/LCF. We anticipate being granted access to the large computational facilities at ORNL/LCF in the near future. Effort is presently underway to arrange such access.
- Per-site facilities requirements to conduct basic research and preliminary tests will be evaluated on a case-by-case basis.
- The SciDAC Outreach Center is providing access to a Subversion server so we can upload and distributed VisIt source code, track downloads, etc. (see Section 12.7.2). Additionally, we will evaluate use of GForge's collaborative project management software for use within VACET in the upcoming period.

11.4 Communication

We anticipate no substantive change in our existing methods for internal Center-wide communication. We anticipate a twice-yearly All-Hands meeting, with the next one to occur in conjunction with IEEE Visualization 2007.

One potential new avenue for communication will be the GForge server at the SciDAC Outreach Center. GForge includes collaborative project management capabilities. We will evaluate use of that capability for coordinated management of VACET projects.

We categorize the Project Management Plan as being part of the general subject of "Communication." While subject to change, our objective will be to submit an updated Project Management Plan on a twice-yearly basis. This frequency will serve to provide regular updates of progress to DOE as well as document ongoing and effective project management oversight across all of VACET's many different activities. The dates of the twice-yearly PMPs will be near the beginning and middle of each fiscal year. Therefore, the next PMPs will be submitted around 1 November 2007, and 1 May 2008.

11.5 Publications

Generally speaking, while any prediction of publications is somewhat imprecise, we can anticipate that, roughly speaking, each stakeholder project is likely to produce one or more publications. Taking into account all stakeholder projects, we can reasonably anticipate on the order of 15-30 publication submissions over the course of the next year or so. These would span a potentially diverse range of topics: results of basic research in visualization and analytics algorithms, applications of techniques to solve domain-specific problems, management reports, web pages, movies and images.

11.6 Outreach, Presentations, Education and Service

- SC2007 Workshop. Joint with the Visualization Institute, we have submitted a proposal for a workshop at SC2007.
- SciDAC 2007 Program Meeting – Tutorial/Workshop. VACET will conduct a one-day workshop in conjunction with the SciDAC 2007 Program meeting in late June 2007 in Boston, MA. The objective will be to provide training in use of VACET's production visual data analysis software infrastructure to current and potential stakeholders as well as other interested parties.
- VACET team members serve in a leadership role for the 2007 prestigious Dagstuhl Seminar on "Scientific Visualization," by invitation only. K. Joy is part of the organizing committee and C. Johnson, C. Hansen, V. Pascucci, C. Silva, have been invited to give a presentation.
- International Summer School on Scientific Visualization 2007. C. Hansen, and V. Pascucci will give a cycle of lectures for a week at the summer school "Ecoles D'Ete" on the topic of "Advanced Methods in Scientific Visualization" 2007.
- IEEE Visualization 2007. VACET team members occupy many positions within the IEEE Visualization 2007 conference team. K. Joy is the conference chair. C. Hansen leads the Technical Program Committee. Bethel is a member of the Technical Program Committee. Most VACET members are Technical Paper reviewers.

12 Per-Project Accomplishments and Plans

The subsections that follow convey accomplishments, plans, and risks on a per-project basis.

12.1 Accelerator/HEP

12.1.1 High Quality, Interactive Rendering of Particle-Based Datasets

VACET Team: S. Parker, Utah (Team Leader) and Thiago Ize, Utah.

Customer/stakeholder: Paul Hamill and Peter Messmer at Tech-X Corporation.

Stakeholder need(s): Tech-X has previously showed the open-source POV Ray ray tracer to produce high quality ray traced visualizations of both fusion and accelerator data. We set out to demonstrate that these images could be rendered at multiple frames per second, for both interactive visualization and high quality publication images. In this context, visual cues such as shadows (including soft shadows), high quality lighting, and even reflections can help convey spatial relationships effectively.

Milestones/deliverables:

- Demonstrate interactive rendering of particles similar to the offline renderings already performed at Tech-X[18] of output from the Vorpal code.
- Incorporate some of the particle binning algorithms so that the heightfield visualizations can be performed interactively as well.
- Install Manta at Tech-X.

Task Dependencies. The tasks are thus far largely independent of one another.

Accomplishments.

We started work on this project in February and we've accomplished all of the original goals. They originally wanted a way to move lights interactively so that they could see how high off the ground their particles were based off of the shadows. In the end it turned out that we could also do many more things for them than they expected. We gave them soft shadows that do an even better job (although it's not yet fast enough for interactivity since it's a temporary solution.) We showed how we can do animations (with interpolation if wanted) in real-time. We showed that we can even handle their other data sets, such as their Vorpal cavity datasets[19], which require "complex geometry". For an example of our work, see Figure 3 below as well as link on the VACET twiki[20].

Currently, the remaining tasks are:

1. User interface, such as for recording camera paths and stopping/starting the animation when the user moves the mouse.
2. Adding particle culling.
3. Faster soft shadows.
4. Implement path tracing, ambient occlusion, and material models for higher quality images.

In the end, Peter seemed happy with the work and said "I think it's obvious that you have all the building blocks in place for our viz needs in VORPAL."

Work Targets Next Six Months.

- Apply Manta to the data produced by their upcoming runs.
- Install Manta at Tech-X.

[18] http://fusion.txcorp.com/∼messmer/vorpal/

[19] http://fusion.txcorp.com/∼messmer/cavities/

[20] see: http://www.vacet.org/gallery/index.html#accelerator

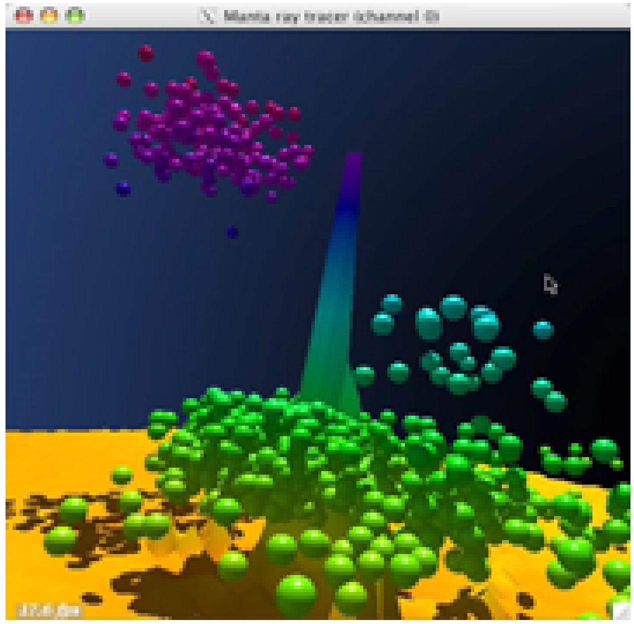

Figure 3: The heightfield represents particle density from a VORPAL dataset of a laser wakefield acceleration. The spheres are individual particles whose velocity exceeds a certain threshold. Since we only had access to a single timestep of data, we randomly perturbed the heightfield and spheres to show we can handle time-varying datasets. T. Ize, SCI. Data data provided by Tech-X corporation and Peter Messmer.

- Incorporate improved material models to further enhance the visual quality of the images.

Next 12-24 months: If they find this tool useful, the primary improvement would be an improved user interface. The python-based user interface in Manta would allow us to create a custom GUI for their applications if necessary.

12.2 Astrophysics

12.2.1 Spectrum Synthesis Visual Analytics

VACET Team Members: Cecilia Aragon, LBNL (Team Leader); Sarah Poon (LBNL), Karl Runge (LBNL)

Stakeholder: Peter Nugent working with Stan Woosley

High Level Stakeholder Needs: Supernova spectral analysis via visual analytics and interactive visualization. See the VACET twiki[21] for more information.

Milestones/Deliverables for Next 6 Months

Start date on this project was January 2007. A1 is ongoing. Expected completion dates.

Task 1. Research and evaluate various approaches to spectral data visualization. Expected completion: 5/31/2007.

[21] http://www.sci.utah.edu/vacetwiki/index.php/Collab:Astrophysics

Task 2. Select best approach(es). Estimated start and completion dates: 6/1/2007, 6/30/2007.

Task 3. Prototype visual analytics tool for spectral analysis and classification. Estimated start and completion dates: 7/1/2007, 9/15/2007.

Task 4. Apply tool to experimental supernova spectral data set. Estimated start and completion dates: 9/16/2007, 11/1/2007.

Accomplishments

Cecilia Aragon began work on this project in January 2007. Work has begun on the first task, which includes scoping the project, including preparation of requirements documents and planning meetings with Peter Nugent and Wes Bethel.

Plans for Next 12 to 24 months:

1. Feature detection and classification via machine learning and clustering algorithms applied to the spectral datasets.
2. Spectra parameter fitting.
3. Temporal analysis of supernova spectra.

Start date for these activities could be as early as Oct 2007. Estimated duration of these three tasks: 12-18 months.

Risks

Task 2. Risks: Inability to define appropriate metrics for evaluating spectral data comparison approaches. Inability to fully evaluate approaches due to difficulties finding appropriate metrics. Inability to meet with customer due to scheduling conflicts. Possible wrong choice of approach. Probability: low.

Task 3. Risk: Chosen approaches might not work well and we would have to start over. Software performance might be insufficient. Likelihood risks: low-moderate.

Task 4. A4: start date Sept 16, 2007; expected duration two weeks. Potential risks: moderate. Risk: Inability to integrate with existing supernova spectral database. Assistance from experts unavailable due to scheduling conflicts. Likelihood: low-moderate.

General The most significant risk is lack of personnel with sufficient expertise in the areas of machine learning and statistical algorithms (LBNL recently lost a staff member with such skills to Google). The remediation for this risk is to hire such staff. Likelihood of the risk occurring is high: we have no such staff right now. The impact is medium: without such staff, progress in the analysis parts of this project will be hampered, although progress in other areas can proceed.

12.3 Climate

12.3.1 Comparative Analysis and Visualization for Climate

VACET Team Members: Claudio Silva, Utah (Team Leader), Marty Cole (Utah), VisTrails team (Utah).

Customer/Stakeholder: Dean Williams and the entire CDAT user community.

Stakeholder Needs We hope to show through example how CDAT users can use their scripts in VisTrails, and get improved productivity. In particular, improved data and process provenance, improved user interfaces for parameter studies, etc. The first target is a script from the CDAT documentation. We expect that the list of needs evolves as evaluation of preliminary work progresses, and through collaboration with the CDAT development team.

Milestones/deliverables

Task 1. Integrate installation requirements in python for both CDAT and VisTrails.

Task 2. For the selected test script, provide wrapped modules in the CDAT VisTrails package, and display the resulting image in the VisTrails spreadsheet.

Task 3. Get feedback from Dean Williams and his team.

Task 4. Expand the wrapped module set based on feedback.

Task 5. Possible tighter integration between spreadsheet and the vcs.Canvas (currently file based).

Subtasks. Marty Cole to work on the list of tasks. Other members of the VisTrails development team have been working on improving functionality to make his work easier, including easier installation on Linux, and other platforms, improved spreadsheet functionality, and other enhancements to the system to make it easier to integrate.

Task Dependencies.

Task 2. Requires completion of Task 1.

Task 3. Requires completion of Task 2.

Tasks 4 and 5. Requires completion of Task 3.

Schedule Estimated duration, start date, completion date: started work Nov 06, Duration and completion depend on feedback.

Accomplishments

Resulting image viewed inside the VisTrails spreadsheet. Enlarge Resulting image viewed inside the VisTrails spreadsheet. The pipeline of wrapped CDAT interface modules that produced the image in the spreadsheet. Enlarge The pipeline of wrapped CDAT interface modules that produced the image in the spreadsheet.

Task 1. Completed.

Task 2. Completed. The package has been delivered to Williams for evaluation. See Figure 4.

Task 3. In progress.

Work targets for the next six months

1. Get feedback from Dean and his team.
2. Identify best path forward to wrap CDAT interface.
3. Determine if a tighter integration between the canvas and the spreadsheet is a priority or not.
4. Identify a target to show parameter exploration capabilites.
5. Support any CDAT users who wish to integrate their scripts.

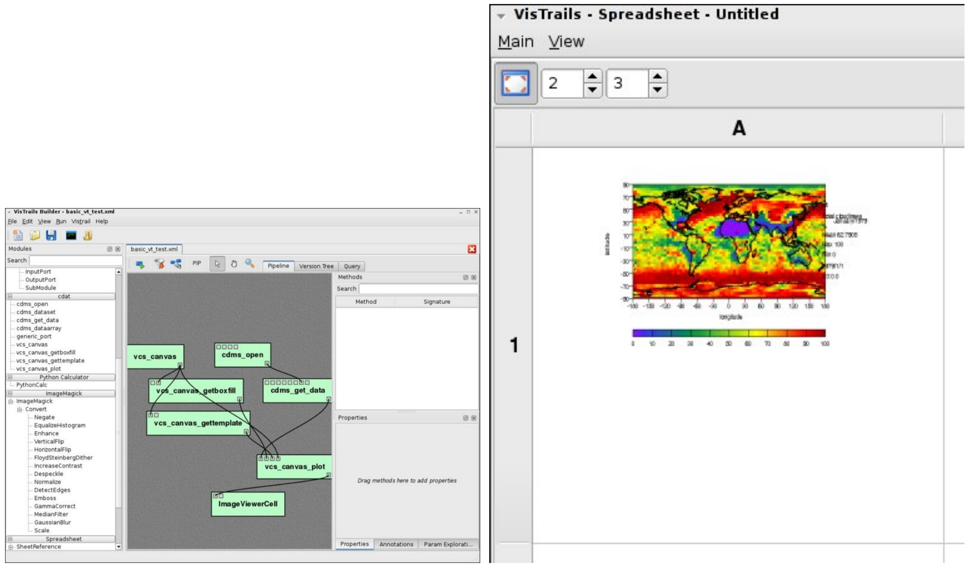

Figure 4: The image on the left shows the VisTrails interface in which portions of a CDAT processing pipeline are assembled into a dataflow-like presentation to produce an interactive application. The results of this particular processing pipeline are shown in the image on the right, which is the visual output of the VisTrails pipeline execution. This is a significant accomplishment because it demonstrates: (1) effective integration of CDAT into a visual programming environment (VisTrails); (2) once integrated, then the capabilities of VisTrails can be brought to bear on CDAT-based climate data analysis problems. More information about VisTrails is located here: http://vistrails.sci.utah.edu/index.php/Main_Page.

Next 12-24 months: At this point in time, it is no practical to plan that far ahead without Task 3 completion.

12.3.2 Advanced Production-Quality Visual Data Analysis for Climate

VACET is committed to supporting the needs of the Community Climate System Model (CCSM) Consortium[22] in collaboration with the Earth System Grid[23].

Mission. The most advanced climate modeling systems seek to enable a new deeper understanding of the dynamics of global carbon cycle, atmospheric chemistry, land and ocean ecological processes and their coupling with climate. This will allow pursuing reliable answers to fundamental questions related to climate variability and global change at time scales ranging from decades to centuries. In this effort VACET will work in close collaboration with the Earth Science Grid and provide new advanced data analysis and visualization tools to the CCSM Consortium and the climate modeling community in general. One target will be the deployment of a first set of tools by FY08, in time to facilitate the analysis of data for the 5th assessment report of the Intergovernmental Panel on Climate Change (IPCC)[24].

VACET Team Members:

- Valerio Pascucci, LLNL (Team Leader).

[22]http://www.cgd.ucar.edu/csm
[23]http://www.earthsystemgrid.org
[24]http://www.ipcc.ch

- Peer-Timo Bremer (LLNL) core visualization techniques, data comparison, topology.
- Marty Cole (Utah) deployment of visualization libraries.
- Jamison Daniel (ORNL) user interaction, case studies, testing of tools.
- Ming Jiang (LLNL) topological analysis, VTK components.
- Daniel Laney (LLNL) data management and streaming.
- Ajith Mascarenhas (LLNL) feature tracking.
- Xavier Tricoche (Utah) 2D topological visualization and analysis.

Customer/stakeholder:
Dean Williams, LLNL, Earth Systems Grid. John Drake, LLNL. Community Climate System Model Consortium (CCSM). Intergovernmental Panel on Climate Change (IPCC).

Stakeholder need(s):
A. Deploying advanced visualization capabilities into the CDAT tool and create a clear path for similar integration in other tools.

B. Extend the visualization software to incorporate domain specific requirements, data formats, and vector field visualization.

C. Support time-dependent and cross-dataset comparison, visualization and analysis.

D. Develop new analytic capabilities for climate data (first deployed into CDAT/VCDAT).

E. Develop a visualization and data analysis scenario for understanding of complex coupled phenomena such as the multi-scale dynamics the complete carbon cycle on earth.

Milestones/Deliverables to Meet Each of Above Needs

Deploying Advanced Visualization Capabilities
The customers would like to integrate state of the art and new 3D visualization capabilities into the CDAT/VCDAT package for rendering of climate data. This need will be addressed by providing a Python integration for a visualization library based on a streaming data model, which addressed the main request of having a light weight and portable external package to be released with CDAT and easy to integrate with other deployment targets. The work will proceed in three steps:

- Defining a high-level interface into ViSUS.
- Creating a Python wrapper around the interface.
- Integrate the interface into both CDAT (the command-line tool) and VCDAT (the GUI version).
- The final deliverable for this stage is a 2D/3D contouring and 3D volume rendering interface for (V)CDAT for regular (Cartesian) grid data.

Customized Extensions
Climate researchers use a variety of special data formats and have some domain specific requirements not common in generic scientific visualization tools. In particular, we plan to provide visualization with data projected onto a model of the earth (sphere or geoid), enhance the 3D visualization with underlying geospatial information (i.e. satellite images or boundaries of the nations), and to natively handle various types of meshes used in the climate community (e.g. geodesic grids, curvi-linear grids, etc.). Furthermore, we plan to address a need for vector field visualization tools, for example, using streamlines and/or LIC based techniques. The deliverables include extending

the visualization packages with the specialized techniques needed and the development of vector field visualization package.

Multi-dataset/time-dependant Visualization and Analysis

One important aspect of climate research is the constant need to compare and analyze data from different models or from the same model at different points in time. To support this need the visualization should be extended to incorporate multiple linked views, and interactive exploration of (potentially long) time series. The deliverables in this area include the computational infrastructure to handle large time-dependent data sets in (V)CDAT and a user interface supporting cross-model and cross-time comparison and exploration.

Advanced Analytics

The customers have a strong interest in new analytic capabilities that are integrated into (V)CDAT. Of particular interest are novel techniques for defining, analyzing, and tracking features as well as new multi-scale data comparison metrics. To this end we will aim at integrating large scale statistical analysis as well as topology based techniques into CDAT and provide high level interfaces for their usage. Furthermore, we are developing new data comparison metrics which will provide multi-scale structural metrics. There is also a need for new topological analysis based on robust combinatorial data processing extending recent work developed at LLNL for conservative vector fields. This includes finding critical point, creating the topological skeleton, and robustly simplifying vector field topology. The deliverables in this area include a suite of multi-scale analytical tools both vector and scalar data to be integrated into (V)CDAT.

Case Studies: Coupling Dynamic Models and Multiscale Dynamics

The ultimate goal of our work is to provide visualizations and analysis techniques that enable a new deeper understanding of the dynamics of climate models. We will include in our activities direct support for the analysis of particularly complex and interesting case studies. On probable target will be the study and understanding of the global carbon cycle, including atmospheric chemistry, integration of land and ocean ecological processes, and their coupling with climate. This requires data analysis and visualization incorporating several models to answer fundamental questions related to climate variability and global change at time scales ranging from decades to centuries. This is a major challenge requiring close collaboration with the climate researchers.

(A) Advanced Visualization: Subtasks/milestones/deliverables (37 weeks)

Task A1. Define and Wrap Interface to Contouring Library (8 weeks)

- Define low-level interface: select the appropriate C-library functions (1 day).
- Create typemaps for SWIG to interface Pythons Numeric Arrays with C-data structures (1 week).
- Test typemaps and integrate robust type-checking and useful exceptions to speed-up later debugging (1 week).
- Create contouring Python module regression tests (1 week).
- Implement stand-alone C-based OpenGL routines for performance reasons and to avoid PyOpenGL problems (see risks) (1 week). Shading modes to test evaluate: flat shading, smooth shading, wireframe.
- Create stand alone Tkinter-based sandbox for testing and later integration (2 week).
- Regression testing the code by comparing with the native ViSUS package, performance check (1 week).
- Develop initial user-interface to steer contour visualization (trackball, iso-value changes, File IO etc.) (1 weeks).

Task A2. Define and wrap interface to a basic volume rendering library (9 weeks).

- Define low-level interface (1 day).
- Create typemaps (2.5 days).
- Test typemaps and implement type checking (2.5 days).
- Create volume rendering Python module (1 week).
- Simple regression tests (1 week).
- Create alternative compilation targets for use of 2D texture, 3D texture, and mesa software rendering (2 weeks).
- Implement C-based OpenGL routines and interface the texture based based rendering into PyOpenGL (1 week).
- Integrate the Python module into the sandbox (3 days).
- Regression test, validation, performance analysis (1 week).
- Extent the user interface of the sandbox to incorporate volume rendering (1 weeks).

Task A3. Non-orthogonal, Progressive Data Slicing (9 weeks).

- Extract slicing library from the ViSUS data access module (2 weeks).
- Define low level interface (1 day).
- Create and test typemaps (2 days).
- Create slicing Python module (1 week).
- Regression testing (1 week).
- Implement C-based OpenGL routines for fast rendering (2 days).
- Integrate slicing module into the sandbox (3 days).
- Regression testing performance analysis (1 week).
- Implement prototypical .idx file-format conversion at load time (1 week).
- Performance analysis of idx conversion viability assessment (1 week).
- Prototypical progressive display and viability evaluation (1 week).
- Extend user interface of the sandbox to incorporate non-orthogonal slicing (1 weeks).

Task A4. Prototypical CDAT integration restricted to one target platform (2.5 weeks).

- Integrate the sandbox into CDAT by hand on some selected machines (e.g. a presentation laptop) (1 week).
- Evaluate the user-interface with Dean's team decide on necessary changes and additions (2 days).
- Discuss the need/choice of colormap/transfer function editor with Dean and/or "standard" color maps / transfer functions (1 week).

Task A5. Interface design, integration and adaptation (3 weeks).
Task A6. Incorporate/implement colormap/transfer function editor (3.5 weeks).

- Colormap editor (1 week).
- Transfer function editor (2 week).
- Develop a small library of preexisting transfer functions and allow to grow the database (3 days).
- Test integration of transfer function design with contour spectrum and other histograms (1 week).

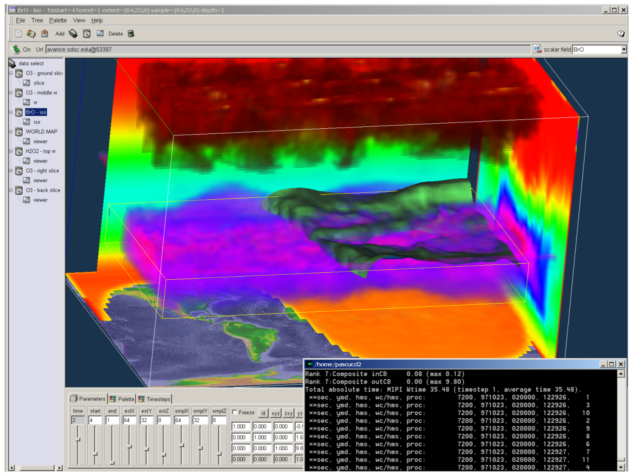

Figure 5: Preliminary test of a climate modeling dataset loaded within the ViSUS data streaming and rendering framework.

Task A7. Final (V)CDAT integration (5 weeks).

- Collect all necessary libraries into the CDAT distribution (2 days).
- Adapting the build system / uncouple libraries (2 weeks).
- Compatibility testing for Windows, Linux, and MAC platforms (4 weeks).

Task Dependencies

The tasks and subtasks are mostly sequential.

Accomplishments

(A) Preliminary study of the CDAT environment and developed a detailed plan for the integration of the 3D visualization capabilities (in collaboration with Dean Williams). Started wrapping the contouring library in Python for integration in (V)CDAT (see Figure 5). C: Prototypical implementation of a univariate multi-scale topological error metric.

12 month:

- Deployment of basic 3D visualization package in (V)CDAT
- Determination of the most desired extensions (vector fields, multi-data set, time-dependent) and initial results
- Research prototype for extended analytic capabilities

Work targets for the next three to six months.

- Working Python interface to the contouring library (A1).
- Initial sand box implementation (A2).
- Working Python interface to the volume rendering library (A2).
- Python module for contouring, volume rendering, and slicing (A2, A3).
- User feedback leading to final interface design (A5).

Risks and Opportunities

1. **ViSUS code unreliable.** Likelihood: low. Impact on project: high. Remediation: Debug/rewrite. However, the ViSUS libraries have been thoroughly tested and a mission-critical bug is unlikely. Team members are familiar with the code and eventual unexpected debugging will not be too time consuming. Potentially, the volume rendering could be replaced with the SCIRun library (which might be a later target anyway due to its potential for higher Image quality).
2. **OpenGL/Python integration** The current OpenGL extension to TkInter called Togl is no longer supported and has some known compilation problems. Likelihood: low. Impact on project: medium. Remediation: We are expected to use Togl only in a very minimalistic way in which no problems have occurred in the past. Willianms' team has extensive knowledge of python and seem well able to fix most foreseeable problems. The worst case scenario is to move toward a standalone GUI started from but not integrated into CDAT (i.e. a separate main loop).
3. **Moving to NumPy.** At some point the CDAT project will move from Numeric.arrays to NumPy which will require a change of all wrapping code. Impact on project: high. Remediation: Depending on when this move happens, it will likely require at least one-two weeks to switch the code base. There is a small possibility that necessary functionality will be missing from NumPy and a more extensive re-write might be necessary. Looking at the current state of Numeric and NumPy this seems unlikely. To avoid major problems we will use only constructors that are already used in CDAT so that we can adopt fixes developed in the ESG group.
4. **Volume Rendering compatibility problems.** The current implementation of the volume rendering library relies on some OpenGL features (3D texture mapping) for higher performance. These features may not be present on all necessary platforms. If that is the case we will need to provide alternative compilation targets that fold automatically back to the basic slower code based on 2D texture maps. Likelihood: medium-low. Impact on project: high. Remediation: Ignoring the performance issue the render can be extended to include a software-only version based on Mesa. Another option is to switch entirely to the SCIRun volume render.
5. **Transfer function design problems.** The design of good transfer functions is non-trivial even for experts and it is unclear whether there exists a "good" editor for general users. Creating an editor might therefore be less a problem of time investment but rather of developing a novel interface paradigm. The transfer function is an integral part of any volume rendering and without it volume rendering is unlikely to be adopted in the climate community. Likelihood: medium. Impact on project: medium (with a critical impact on the volume rendering aspect). Remediation: There exists research on automated transfer function design that could be applied with some time investment. Furthermore, it might be possible to write a meta-editor which adapts pre-existing transfer functions to a limited set of problems. A large library of pre-defined transfer functions would also alleviate the problem. Finally, a well designed tutorial could be used to educate CDAT users in the use of one or multiple of the advanced transfer function editors developed within the visualization community (e.g. multi-dimensional transfer functions).
6. **CDAT integration problems.** CDAT is required to build on a large variety of platforms some of which might be drastically different from "standard" visualization systems. Potentially there exist platforms which miss crucial components (e.g. OpenGL drivers) which cannot be included into the CDAT distribution. Likelihood: very low (we do not expect to be dependent on any advanced OpenGL feature and we will run the code on low end platforms

such as a laptop). Impact: low (high impact only for a few users). Remediation: We will make sure that our code compiles on the systems of our direct costumers. Other users in the general community will need to turn off our module or upgrade their system.

(B) Customized Extensions (51 weeks).

Task B1. Integrate geospatial information (14 weeks).

- Create file reader for the image formats used in CDAT (2 weeks).
- Wrap NGA code for high resolution imagery and KML data (2 weeks).
- Implement composite rendering superimposing contours/volume rendering over maps (3 weeks).

- Avoid storing multiple data copies.
- Automatic registration between 2D and 3D data.
- Integrate two glut main loops / fix any issue with interfering global rendering variables/structures.

- Integrate code into the sandbox (2 weeks).
- Regression testing (1 week).
- Porting code to selected CDAT platforms and integrate into CDAT build system (4 weeks).

Task B2. Integrate terrain visualization (9 weeks).

- Implement SOAR index mapping (2 weeks).
- Explore SOAR-based data re-ordering for performance [this is similar to the idx conversion under A3] (2 weeks).
- Integrate code into the sandbox (2 weeks).
- Regression testing (1 week).
- Porting code to CDAT platforms integrate into build system (2 weeks).

Task B3. Offline-rendering and movie creation (17 weeks).

- Develop file formats and readers/writers to save and load scenes (XML parser for saving/,loading view point, color map, interface configuration, etc.) (2 weeks).
- Incorporate off-line rendering into a virtual window of any arbitrary resolution (i.e. "Print" button) (4 weeks).
- Off-line rendering into compressed image formats (e.g. png) (1 week).
- Create movie pipeline for static view-points (2 weeks).
- Create movie pipeline for dynamic view-points (i.e. fly-through) (3 weeks).
- Parallelize movie creation for selected batch systems (2 weeks).
- Regression testing (3 weeks).

Task B4. General embedding of regular domains : Allow regular grids to be warped by a general xyz mapping (17 weeks).

- Adapt iso-contouring code to warped grids (2 weeks).
- Regression testing (2 weeks).
- Adapt volume rendering to warped grids (8 weeks).
- Regression testing (3 weeks).

Task B5. Tetrahedral and general meshes (4-6 months).

- Decide which are the important grid types that should be supported (1 week).

- Create file readers (2 weeks).
- Implement re-sampling procedures to deal with grids that cannot easily be mapped to the visualization algorithms (4 weeks).
- Adapt visualization tools to the remaining grid types (2-4 months).

Task B6. Visualization of computational domains for debugging (6 weeks).

- Implement high-performance OpenGL routines to visualize the underlying grid structure (rather than the data) (1 week).
- Rendering/shading modes: wireframe, hidden lines.
- Imploded cells.
- Integrate into sandbox (3 days).
- Develop query tools and visual aids (3 weeks).
- Highlighting, picking, range queries, clipping planes.
- Porting to CDAT platforms (2 weeks).

Task B7. Vector field visualization (4-6 months).

- Implement/adapt standard vector field visualizations such as streamlines and LIC based visualization for 2D gridded data (1-2 months).
- 3D streamline visualization of regular grid data (1-2 month).
- Visualizing derived quantities such as vortices, shockwaves, etc (1-2 month).
- Integrate visualization code into the sand box (3 weeks).
- Regression testing (2 weeks).
- Porting to CDAT platforms (2 weeks).

Task Dependencies

The tasks and subtasks are mostly sequential.

Accomplishments

Work has not yet commenced on this family of tasks.

Work targets for the next three to six months.

We do not expect to begin work on these tasks in the next three to six months.

Risks and Opportunities

1. **Volume rendering for warped/general meshes.** Adapting the volume rendering to non-cartesian meshes is highly non-trivial and likely to be computational expensive since it requires sorting the data and error prone due to many special cases. Likelihood: high. Impact on project: medium (high for the volume rendering aspect). Remediation: We have some code available that implements a general software renderer for tetrahedral meshes. Other options are to use better code that might be available in Utah or regrid the data.
2. **Unusual grids.** The climate community might use grids very different and non-adaptable to visualization. Likelihood: medium. Impact on project: low. Remediation: One can always perform a resampling on either regular or tetrahedral meshes.
3. **Opportunity: unstructured tet meshes may not be needed.** Likelihood: unknown. Impact on project: low. Remediation: We will adjust out plan and achieve our targets sooner.

4. **User interface for vector fields.** Vector field visualizations depend on many parameters such as streamline density and location, integration step size, etc.. It is not clear how these can be exposed in a meaningful way. Likelihood: medium. Impact on project: low. Remediation: More work on the interface and if necessary some trial and error session with the users to determine useful defaults.
5. **Performance problems dues to the extra cost of grid mapping and/or multiple rendering threads.** If used in full there will be up to 4 rendering threads running at any one time (iso-surfaces, volume rendering, terrain visualization, 2D imagery) which will impact performance. Likelihood: medium. Impact on project: medium. Remediation: Depending on the severity more time can be allocated to optimize the code and/or make better use of modern GPU's if available. Features can always be selectively turned on and off which should allow useful exploration in any case. For some users a parallel rendering framework would be a option and/or remote rendering (see C).

(C) Multi-dataset/time-dependant Visualization and Analysis (8-11 months).

Task C1. Multiple linked views (7 weeks).

- Extend the sandbox interface to allow multiple linked viewports showing multiple data sets (4 weeks).
- Support multiple data sets in memory at the same time.
- Cross data set registration.
- Allow dual/multiple use of interface functionality, e.g. use the same iso-surface slider for all viewports.
- Link and unlink visualization parameters such as iso-value, viewpoint, transfer function, etc.

- Port to CDAT platforms (3 weeks).

Task C2. Support time-dependant visualization (2-4 months).

- Develop data management framework to load/cache time steps for interactive visualization (1-2 months).
- Explore the feasibility of converting the data to visualization-friendly data formats ,i.e idx, (1-2 months).

Task C3. Parallel/remote visualization (4-7 months).

- Spread the visualization tasks over multiple threads to utilize multi-core machines (1-2 months).

- Distribute the visualization across a local network (2-3 months).
- Enable remote streaming visualization utilizing a client-server architecture (1-2 months).

Task Dependencies

The tasks and subtasks are mostly sequential.

Accomplishments

Work has not yet commenced on this family of tasks.

Work targets for the next three to six months.

We do not expect to begin work on these tasks in the next three to six months.

Risks and Opportunities

1. **Code compatibility problems.** Cross-platform parallel code is very difficult to develop and might be too ambitious. Likelihood: high. Impact on project: low (most users are not expected to have easy access to parallel resources). Remediation: Concentrate on few common platforms in particular those at visualization centers (LBNL, LLNL, ORNL) and thread based implementations. Jump directly to remote visualization using shared resources.
2. **Data size.** Depending on the common data sizes single desktop, interactive, time-dependent visualization might simply not be feasible. Likelihood: medium. Impact on project: low. Remediation: Remote visualization and a bigger focus on hierarchical techniques.
3. **Remote visualization problems.** There may be problems with firewalls when dealing with remote data streaming. Likelihood: high. Impact on project: medium. Remediation: Worst case: remote visualization will be enabled only within selected networks. Expected solution: use of ssh tunnels for all "difficult" situations.

(D) Advanced analytics (23-25 months + vector field research)

Task D1. Integrate 2D structural analysis into CDAT (3 months).

- Extract a library from topology applications (3 weeks).
- Wrap the low-level library (2 weeks).
- Integrate into sandbox (2 weeks).
- Topology interface (critical points, manifolds, cells).
- Topology based highlighting, picking.
- Regression testing (2 weeks).
- Port to CDAT platforms (2 weeks).
- Interface evaluation (1 week).
- Create high-level command-line interface for CDAT (2 weeks).
- Create high-level user-interface for VCDAT (1 week).

Task D2. Flexible feature definitions: Create an interface allowing the user to create feature definitions on the fly (5 weeks).

Task D3. Robust feature tracking (7 months).

- Integrate merge tree code with flexible feature definitions (4 weeks).
- Develop hierarchical feature tracking (3 months) [combustion project].
- Specialize topological tracking to climate related problems (1 months).
- Porting to CDAT platforms (4 weeks).
- Interface/visualization (1 month).

Task D4. Large-scale statistics (6-8 months).

- Develop fast hierarchical methods to access data and collect statistics for large data (2-3 months).
- Create portable library and Python wrapper (4-6 weeks).
- Integrate into sandbox (2-4 weeks).
- Evaluate practical usability and impact (1 week).
- Specialize tools and interface for climate analysis and adapt the interface (1-2 months).
- Porting to CDAT platforms integrate build system (1 month).

Task D5. Multi-scale data comparison (6-10 months).

- Develop mathematical framework (1-2 months).
- Extend code base (e.g. topology code, statistics library) (1-2 month).
- Develop and evaluate different metrics (2-3 months).
- Interface/visualization (2-4 weeks).
- User evaluation (1 week).
- Porting to CDAT platforms integrating build system (2-4 weeks).
- Specialize interface visualization fro climate analysis (4-6 weeks).

Task D6. Vector field topology (15-24 months).

- Robust definition of critical points (1 month).
- Developing "discrete" 2D vector field theory (4-5 month).
- Robust computation of 2D vector field topology (2-3 month).
- Defining topological simplification, error metrics, and hierarchy (2-4 month).
- Extension to 3D vector field topology (6-11 month).

Task Dependencies

The tasks and subtasks are mostly sequential.

Accomplishments

Work has not yet commenced on this family of tasks.

Work targets for the next three to six months.

We do not expect to begin work on these tasks in the next three to six months.

Risks and Opportunities

1. **Opportunity: in stage D3 we may be able to unify and leverage the work in support of combustion simulations.**. Likelihood: high. Impact on project: high. Remediation: adjust the plan and deliver the tools earlier than planned.
2. **Research success uncertain.** In general, many of the techniques depend on additional fundamental research that might not pan out or take significantly longer. Likelihood: unknown. Impact on project: medium. Remediation: Fall-back to traditional techniques.
3. **Flexible feature definition elusive.** A general framework that allows random data access during the feature definition can become a severe performance problem. Likelihood: medium. Impact on project: low. Remediation: Restrict the feature definitions to filters of certain width, streaming filters, and/or integrate progressive framework.
4. **Feature tracking issues.** Features in climate models might be too complex to be defined in closed mathematical terms. Likelihood: low. Impact on project: medium. Remediation: Multi-layered feature definition on derived quantities. For example, allow features to be defined using a combination of multiple input fields in combination with derived quantities (vorticity etc.).
5. **Large scale statistics.** This combination with large scale data access and statistical methods is new and will require close collaboration of the LLNL team with the ORNL team on the development of a working prototype. The novelty increases the risk of delaying the deployment. Likelihood: medium. Impact on project: low. Remediation: CDAT already carries an interface to the "R" statistics package from which one can reuse or adapt components if specialized techniques fail.

6. **Multi-scale data comparison.** The ideas underlying this aspect of the plan have not yet been tested and probably require some more research. Likelihood: medium. Impact on project: low. Remediation: The exists a large array of standard comparison techniques that can be implemented if the novel specialized metrics fail.
7. **Vector field topology.** The theory for robust/combinatorial vector field topology has not been developed yet. Experience in the scalar case suggests that there will be a significant ramp-up time but also great benefits. Likelihood: unknown. Impact on project: low (vector field topology is only a small specialized aspect). Remediation: Several VACET members have experience in numerical methods for computation of Vector field topology (especially Xavier Tricoche and Ming Jiang) and their results could be used as an intermediate/substitute solution.

(E) Case studies based on coupling of multiple models and multi-scale dynamics.

Task E1. Work with the users to develop a high-impact case study based only on visualization tools (2 months).

Task E2. Introduce analytical component and quantitative feedback complementing the visualization (2 months).

Task E3. Refine and integrate the visualization and analytics components, develop conclusions and document results with images, videos and writing a technical report (4-6 months).

Task Dependencies

The tasks and subtasks are mostly sequential.

Accomplishments

Work has not yet commenced on this family of tasks.

Work targets for the next three to six months.

We do not expect to begin work on these tasks in the next three to six months.

12.4 Combustion

The VACET team is performing R&D on several projects aimed at providing new visual data understanding capabilities to the combustion research community.

12.4.1 Advanced AMR-Based Temporal Analysis and Visualization

VACET Team Members: Hank Childs, LLNL (Team Leader); Gunther Weber, LBNL.

Customer/stakeholder: Marc Day, LBNL, of the CCSE group under John Bell.

Stakeholder need(s). Deploy capability to better understand underlying processes of combustion calculations. In this case, Marc would like to be able to identify a surface of interest (a contour that represents the start of the flame) and be able to walk backwards in time using a vector (e.g. use the velocity to determine where the chemicals currently on the flame surface were located at the previous time slice), and then repeat all the way back to the beginning of the simulation. As the surface is traced backwards (to create a volume), Marc would like different scalar variables (such as species information) to be evaluated at each time. Once these volume are created, Marc would additionally like to do analysis on these volumes, by calculating ratios using combinations of surface areas on the surface, volumes of the swept out prisms, and the scalar fields on the mesh.

Subtasks/milestones/deliverables.

Task 1. Create a VisIt operator that will start with an initial surface and iterate backwards through time, generating the swept volume as it goes.

Task 2.

Establish a file format that can capture this volume without exceeding available memory. That will likely involve dumping the results from each time slice as it goes.

Task 3. Create a VisIt operator that will modify the contour to no longer contain extremely small triangles. These triangles create numerical problems. "Small" is measured with surface area and high aspect triangles must also be removed.

Task 4. Add routines to analyze the resulting files and create the 1D curves Marc is truly interested in.

Task Dependencies Dependencies: Tasks 1 and 2 will be performed concurrently. Task 3 is independent of 1 and 2. Task 4 depends on 1, 2, and 3.

Estimated duration: Tasks 1 & 2: 1 week. Task 3: 1 day. Task 4: 1 week.

Accomplishments

Hank Childs implemented most of Task 1. The current implementation is able to take a surface of interest and "animate backwards", but several points along the surface don't move, even though the vector field there is non-zero.

Work targets for the next six months.

Complete Tasks 1, 2, 3 and 4. Hank Childs will be doing the work with Weber observing and providing support.

Risks

1. VisIt infrastructure for parallel point location is buggy. Likelihood: low. Impact on project: medium. Remediation: would have to debug point location routines in VisIt (this could take a while). Affected: Task 1.
2. Creating "acceptable" triangle mesh contains hidden costs. Likelihood: low. Impact on project: high. Remediation: should iterate with Marc to better understand what will and will not work. Marc suggested using QSlim, but that is a heavyweight project (although it caps the cost). Affected: Task 3.
3. Analysis of resulting volume may go beyond the ratios that Marc initially mentioned. Likelihood: high. Impact on project: medium. Remediation: I don't think there is a remediation here. The goal is not to "get in and get out". It is to help Marc solve science problems. If this requires more development than initially foreseen, then that's what it will take. However, the likelihood that this becomes an extremely large time investment (i.e. multiple man-months) is extremely low. Affected: Task A4.

Timeline Next Six Months

Start date: 4/6/07. Targeted end date: 5/15/07.

Next 12-24 months. This project will complete in the next six months.

12.4.2 Topological Combustion Data Analysis – Part I

VACET Team.

- Valerio Pascucci, LLNL (Team Leader).
- Peer-Timo Bremer (LLNL) topological feature extraction.
- Daniel Laney (LLNL) data management and streaming.
- Ajith Mascarenhas (LLNL) feature tracking.

Customer/Stakeholder(s). Jacqueline Chen, SNL.

Customer/Stakeholder need(s). Topological characterication of features in a combustion process such as extintion pockets. Development of a robust tracking scheme and practical use for relialable quantitative analysis. These capabilities offer traction on the problem of visual and quantitative analysis, something very much needed as part of the scientific data understanding process for combustion research. These capabilities are not provided by any other source, particularly not in scalable visualization tools (production or research).

Deliverables to meet Stakeholder Needs. Deliver topological analysis software to aid understanding of combustion process. Jacqueline Chen's group works with large 2D and 3D time-varying combustion simulation datasets, and needs analysis tools to aid their understanding of the combustion process. They are interested in analyzing the data to detect the dominant re-ignition mechanism of extinction pockets.

Task 1. Identify data formats, develop data read/write modules, format interchange modules, assess time to marshal data (2-3 weeks).

- Develop read/write software from Fortran Binary / NetCDF to Brick of Value (BOV) for VisIt and IDX for ViSUS (1 week per format).
- Test software. (3 days per format).
- Data movement and preprocessing
    1. Expected data size and processing steps to prepare data for analysis: (1) current "Excite Jet" 3D data set is 43 GB per time step (242time steps), stored as a single file per time step on HPSS tape; (2) the expected preprocessing pipeline will be: (a) Download one timestep from tape to disk (either NFS or Lustre): 1 hour minimum (assuming available disk space); (b) extract required fields into the proper format for our tool chain 15 minutes per field per time step; (c) Store these files back to tape and transfer to LLNL to mitigate risk 1 minutes (local HPSS), 90 minutes LLNL
    2. Estimated time to preprocess the "Excite Jet" data set: (a) 250 hours (10 days, extreme minimum) to download files from tape; (b) 60 hours (3 days, extreme minimum) to extract 1 field from tape, interleved with previous bullet; (c) 370 hours (16 days, extreme minimum, in parallel with above) to move extracted fields to LLNL filesystem.
    3. Quality assurance of data by performing sanity check of value ranges. (E.g. all pressure values must be within acceptable ranges.) (3 days.)

Task 2. Develop basic visualization software for preliminary study and visual assessment of the preprocessed data. (4 weeks.)

Task 3. Develop feature segmentation and tracking software for time-varying 2D combustion data. (30 weeks. Completed: 19 weeks.)

Feature Segmentation

- Develop 2D feature segmentation software using 2D Morse-Smale complexes. (3 weeks.) (Done)
- Test 2D feature segmentation. (2 weeks.) (Done).
- Present preliminary results to analyze a representative small data-set to stakeholder for feedback. (1 day).
- Iterate with Incorporate changes and add functionality as per feedback. (4 weeks.)

Feature tracking

- Develop algorithm to track selected features over time. (2 weeks.) (Done).
- Develop library to unify geometric ID's across pre-processing, feature extraction, and tracking tools (1 week) (Done)
- Implement preliminary version of algorithm. (4 weeks.) (Done).
- Test tracking algorithm. (2 weeks.) (Done).
- Write proof-of-concept software to visualize and interact with segmentation and tracking data. (3 weeks.) (Completed 1 week of effort).
- Present visualization and tracking software to stakeholder for feedback. (1 day.).
- Develop software into full-fledged tool with controls for (Done)
  - Refining segmentation. (1 week.)
  - Sensitivity of tracking. (1 week.)
  - Simplifying tracking data. (1 week.)
  - Presentation at various scales. (1 week.)
- Test software: correctness, robustness, memory usage. (2 weeks.)
- Present software to stakeholder and get feedback. Iterate if necessary. (1 day.)

Task 4. Develop feature segmentation and tracking software for full-scale 3D combustion data. (Total: 70 weeks.)

The design of the software developed in this stage will rely heavily on the lessons learned in stage A3 and will be crafted to integrate it into the visualization and interaction software developed for the 2D case. Special emphasis will be placed on handling large data-sizes.

Feature Segmentation

- Develop strategy to handle increase in data-size due to increase in dimension.
  - Develop streaming algorithm for 3D Morse-Smale complex construction. (4 weeks.)
  - Develop parallel algorithm for 3D Morse-Smale complex construction. (4 weeks.)
- Implement streaming algorithm for 3D Morse-Smale complex construction. (8 weeks.)
- Test streaming algorithm for 3D Morse-Smale complex construction. (8 weeks.)
- Implement parallel algorithm for 3D Morse-Smale complex construction. (8 weeks.)
- Test parallel algorithm for 3D Morse-Smale complex construction. (8 weeks.)
- Present results and obtain feedback from stakeholder. (1 day.)
- Choose best of streaming and parallel strategies and refine software and develop to deployment stage. (4 weeks.)

Feature Tracking

- Develop strategy to handle increase in data-size due to increase in dimension.
  - Develop streaming algorithm for feature tracking. (3 weeks.)
  - Develop parallel algorithm for feature tracking. (3 weeks.)
- Implement streaming algorithm for feature tracking. (4 weeks.)

- Test streaming algorithm for feature tracking. (4 weeks.)
- Implement parallel algorithm for feature tracking. (4 weeks.)
- Test parallel algorithm for feature tracking. (4 weeks.)
- Present results and obtain feedback from stakeholder. (1 day.)
- Choose best of streaming and parallel strategies and refine software and develop to deployment stage.
  - Integrate 3D segmentation and tracking into visualization and interaction software already developed in stage A4. (4 weeks.)
  - Present software to stakeholder and get feedback. Iterate if necessary. (1 day.)

Task 5. Develop critical point tracking software using Jacobi curves for time-varying 2D data. (Total: 15 weeks. Completed: 7 weeks.)

This component complements the results in Task 4 and provides the necessary degreee of robustness.

- Identify appropriate interpolation functions based on data-format. (3 days.)
- As a starting point, evaluate and choose simplest interpolation and develop/adapt algorithm to compute Jacobi curves. (2 weeks.) (Done)
- Implement algorithm. (3 weeks.) (Done)
- Test algorithm on representative data-set. (2 weeks.) (Done)
- Develop visualization and interaction tool. (With proper design we should re-use software of A3) (1 week.)
- Present preliminary results to stakeholder and get feedback (1 day.)
- Develop software to deployment stage based on feedback. We incorporate other interpolation schemes at this stage. (4 weeks.)
- Stress test with available data-sets. (2 weeks.)

Task 6. Develop Jacobi curve computation software for time-varying 3D combustion data. (Total: 20 weeks. Completed: 3 weeks.)

As in the previous stages, going from 2D to 3D requires careful analysis of the impact of data-size increase and strategies to factor in this increase.

- Identify impact of data-size increase on Jacobi curves computation algorithm. (3 days.)
- Develop streaming algorithm to compute Jacobi curves using a small memory footprint. (4 weeks.) (Completed 3 weeks of effort.)
  - Develop test for determining if an edge in the underlying mesh can be finalized.
  - Develop algorithm to classify criticality of finalized edges and to free its memory.
- Implement streaming algorithm. (4 weeks.)
- Test algorithm using representative data-set. (4 weeks.)
- Develop visualization and interaction software. (2 weeks.)
- Present results to stakeholder and obtain feedback. (1 day.)
- Incorporate feedback/new features and iterate (4 weeks.)
- Stress test with full-sized data-set and iterate if needed (2 weeks.)

Task 7. Develop theory for Jacobi curve simplification. (Total: 1-2 years)

This stage is a major undertaking including new important features not available in any tool. Success in this area will have an impact on several topological analysis techniques that use Jacobi curves and lead to software that can be applied in solving real-world science problems that are currently infeasible. We have already a good preliminary understanding of how to approach this problem.

1. Develop definitions of problem in restricted setting of time-varying functions.
2. Understand how cancellation of critical point pairs changes over time.
3. Develop understanding of how change in pairings modifies Jacobi curve structure.
4. Develop algorithm to compute when and how pairing changes occurs.
5. Develop a multi-resolution hierarchy of Jacobi curves based on structural simplification induced by critical point cancellations.
6. Extend theory and algorithms to general setting of pairs of functions on manifolds. In other words, redefine problem in step 1 after removing restrictions on functions and iterate steps 2 through 5.

Task 8. High quality visualizations and movies for publication and distribution.

1. Iterate with scientists to produce high quality visualizations for publication as figures (2 weeks)
2. Plan movies using low quality rendering (2 weeks)
3. Render frames for movies using Povray or other high quality rendering package 50-200 hours per movie.

Task Dependencies

- Stages 2 through 6 depend on 1, but can proceed after development of a read/write module for a representative data format.
- Although 3 and 4 can be performed concurrently, the lessons learned during 3 can speed-up development in 4. Therefore we set 3 to precede 4.
- Similarly 5 and 6 can be performed concurrently, but we set 5 to precede 6.

Accomplishments

In this preliminary stage we have developed some of the fundamental capabilities necessary to engage the costumer in the definition of the characteristics of the features of interest using the formal language of Morse theory. Figure 6 shows the interactive visualization tool used for the exploration of the large combustion datasets even on regular desktop or laptop computers.

- **Task 1**. Meeting: Valerio and Jacqueline Chen discuss requirements, determine potential test data-sets, and data transfer mechanism.
- **Tasks 1 and 2**. In order to facilitate work on later stages, we used 2D raw format combustion data provided by Jacqueline Chen and visualized it using in-house visualization code written using OpenGL.
- **Task 3**. Core functionalities for segmentation implemented by Peer-Timo Bremer, and for tracking implemented by Ajith Mascarenhas. Ajith is currently developing a visualization and interaction tool to analyze a time-varying 2D combustion dataset provided by Jacqueline Chen.
- **Task 4.** Valerio, Timo, and Ajith have developed a strategy to handle time-varying 3D combustion data based on early lessons learned from work on time-varying 2D data.
- **Task 5.** Ajith has developed working code to compute Jacobi curves of time-varying 2D data-sets for two different interpolation functions.
- **Technical Paper**. Paper accepted for publication: "On-line computation of Reeb-graphs: simplicity and speed" to appear in SIGGRAPH 2007.

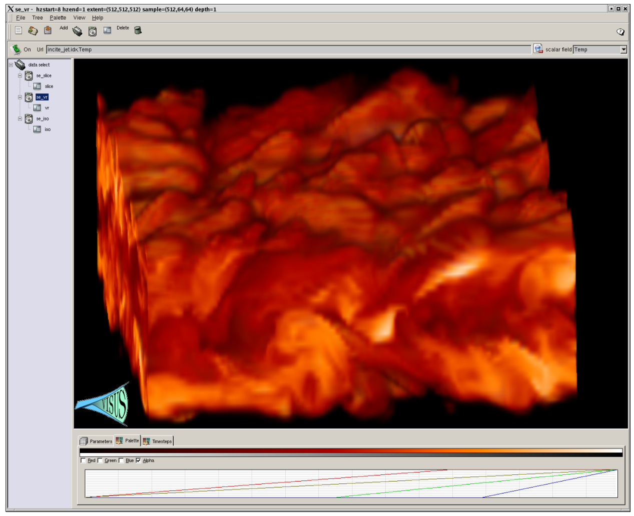

Figure 6: Preliminary test of multiresolution framework applied to a combustion simulation dataset.

Work targets for the next six months.

- Tasks 1 and 2: In progress. Start date: 2/19/07.
- Task 3: In progress. Preliminary version of segmentation and feature tracking software already developed, and will be refined. Expected remaining duration: 11 weeks.
- Task 5: We will build on an existing code-base for Jacobi curve computation for time-varying 2D data-sets using two different interpolation functions. We have finished about 7 weeks of work, and will re-start on 05/15/07. Expected remaining duration: 8 weeks.

Within three months time, finish A3 and demonstrate feature extraction and tracking for time-varying 2D data to stakeholder. Within six months time, finish A5 and demonstrate Jacobi curve computation for time-varying 2D data to stakeholder.

Risks and Opportunities

- **Transition from research prototype code to robust deployment code may be difficult.** Likelihood: high. Impact on project: low. Remediation: We introduce several testing and feedback steps in each stage to ensure that the stakeholder is satisfied with the software. Moreover, all new development effort will use the best software engineering practices and focus on designing modular software. Affected tasks: 2 through 6.
- **Resource scalability: Large resource requirements to handle extremely large data-sizes.** Likelihood: high. Impact on project: moderate. Remediation: We have access to several large high-performance computers in LLNL, LBL, and ORNL that can handle such data-sizes. Affected tasks: 4 and 6.
- **ORNL Lustre filesystem.** We expect the ORNL file system capabilities to be strained by the analysis of 10+ Terabyte data sets Lustre at ORNL is new and untested in a high demand production setting, we expect stability problems and data loss to occur. NFS at ORNL is slower and smaller, requiring manual manual intervention and more data movement to and from tape as analysis progresses. Likelihood: very high. Impact on project: medium, we will plan for these issues. Remediation: moving preprocessed data to LLNL will enable us to spread computation among 6+ file systems and 4+ clusters to mitigate filesystem downtime.

- **Software does not scale appropriate to increase in data-sizes.** Likelihood: high. Impact on project: high. Remediation: We incorporate scalability as a design requirement in each of our tasks. In particular, our plan includes a design of visualization, feature segmentation, and Jacobi curve computation software to handle large data-sizes when moving from 2D to 3D. We will design fast memory-efficient streaming algorithms and also design parallel algorithms to compute 3D Morse-Smale Complex and Jacobi curves. Affected tasks: 4 and 6.
- **Change in requirements from stakeholder.** Likelihood: medium. Impact on project: moderate. Remediation: We add feedback steps in each stage to identify any change in requirements as soon as possible. We keep our software designs sufficiently general and flexible to quickly add more functionality as and when required. Affected tasks: 3 through 6.
- **Jacobi curves are too noisy.** Likelihood: high. Impact on project: moderate. Remediation: We have a practical simplification algorithm that removes noise and produces a coarse-to-fine hierarchy of Jacobi curves. Preliminary studies indicate that this algorithm has potential that we will develop. Stage A7 plans for a full-fledged investigation into the mathematics and algorithmic development based on a thorough understanding of the problem. Affected tasks: 5 and 6.

Next 12-24 months.

- Task 4: Start date: 9/16/07. Expected duration: 70 weeks.
- Task 6: Start date: 6/01/07. Expected duration: 20 weeks.
- Task 7: Start date: 1/01/08. Expected duration: 1-2 years.

12.4.3 Topological Combustion Data Analysis – Part II

VACET Team.

- Valerio Pascucci, LLNL (Team Leader).
- Peer-Timo Bremer (LLNL) topological feature extraction.
- Daniel Laney (LLNL) data management and streaming.
- Ajith Mascarenhas (LLNL) feature tracking.

Customer/stakeholder: John Bell and Marc Day, LBNL.

Stakeholder need(s).

Bell and his team perform computationally based research to model and study the combustion process. Their overall objective is to gain understanding into the physical and chemical processes that make up "combustion."

One of their main current thrust areas is to characterize and understand differences in regions of the domain that are "burning" vs. "non-burning." There are multiple potential approaches to achieve this objective. Currently, Bell's team writes small standalone codes that post-process larger datasets to extract data subsets that are then subjected to analysis. This approach is tedious and time consuming.

In discussions between VACET and Bell's team, we explored the idea of performing topological analysis of time-varying AMR data to first construct a "line skeleton" that represents spatial connectivity of "non-burning" regions over time, then subject that topological structure to rigorous

topological analysis. Both teams felt this approach held promise for helping to shed light on the nature of combustion.

Subtasks/milestones/deliverables.

Deliver topological analysis software that performs structural analysis of combustion features to aid understanding of the combustion process. John Bell's group works with large AMR 3D time-varying combustion simulation datasets, and needs analysis tools to aid their understanding. They will provide us with a boolean volume of "burning" and "non-burning" regions, and we will perform a detailed analysis of the structural characteristics of the network of low temperature regions. To this end we will use topological tools both feature characterization and for robust time tracking.

Task 1. Identify data formats, develop data read/write modules, format interchange modules, assess time to marshal data. (Total: 3-6 weeks.)

- Develop read/write software from Fortran Binary / NetCDF to Brick of Value (BOV) for VisIt and IDX for ViSUS (1 week per format).
- Test software. (3 days per format).
- Data movement and preprocessing. The expected data size and processing steps to prepare data for analysis is as follows:
  - Current data set is 4GB per variable per timestep, with 23 variables per time step (242time steps).
  - The expected preprocessing pipeline will be: 1. Download one timestep from tape to disk (either NFS or Lustre): 1 hour minimum (assuming available disk space); 2. Extract required fields into the proper format for our tool chain 15 minutes per field per time step; 3. Store these files back to tape and transfer to LLNL to mitigate risk 1 minutes (local HPSS), 2.5 hours LLNL.
- Estimated time to preprocess the "Excite Jet" data set:
  - 14 days to download files from tape.
  - 1 week to extract 1 field from tape, interleaved with previous point.
  - 25 days, (in parallel with above) to move extracted fields to LLNL filesystem.
- Quality assurance of data by performing sanity check of value ranges. (E.g. all pressure values must be within acceptable ranges.) (3 days).

Task 2. Develop basic visualization toolkit for preliminary study and sanity check of data. (Total: 2 weeks).

- Adapt or develop software for loading, displaying, and interacting with combustion data.
  - Test loading, display, and interaction using VisIt. (1 week.)
  - Test loading, display, and interaction using ViSUS. (1 week.)

Task 3. Develop topology extraction software for a small subset of the data. (Total: 36 weeks. Completed: 8 weeks.)

Feature segmentation

- Develop 3D Morse-Smale complex code to segment data. (6 weeks.) (Completed 3 weeks.)
- Test segmentation code on subset of dataset. (4 weeks.)
- Develop code to extract ridge-lines from segmentation (4 weeks.) (Completed 2 weeks.)
- Test extraction code on dataset. (2 weeks.)
- Present preliminary results to stakeholders for feedback. (1 day.)
- Incorporate changes and add functionality as per feedback. (4 weeks.)

- Repeat step 4 based on feedback until the software is developed to required state.

Feature tracking

- Develop algorithm to track selected features over time. (2 weeks.)
- Implement preliminary version of algorithm. (3 weeks.)
- Test tracking algorithm. (2 weeks.)
- Write proof-of-concept software to visualize and interact with segmentation and tracking data. (3 weeks.) (Completed 1 week.)
- Present visualization and tracking software to stakeholders for feedback. (1 day.)
- Develop software into full-fledged tool with controls for:)
  - Refining segmentation. (1 week.)
  - Sensitivity of tracking. (1 week.) (Done.)
  - De-noising using persistence filtration. (1 week.)
  - Simplifying tracking data. (1 week.) (Done.)
  - Test software: correctness, robustness, memory usage. (2 weeks.)
  - Present software to stakeholder and get feedback. (1 day.)

Task 4. Develop feature segmentation and tracking software for full-scale 3D combustion data.
Feature Segmentation

- Develop strategy to handle increase in data-size due to increase in dimension.
  - Develop streaming algorithm for 3D Morse-Smale complex construction. (4 weeks.)
  - Develop parallel algorithm for 3D Morse-Smale complex construction. (4 weeks.)
- Implement streaming algorithm for 3D Morse-Smale complex construction. (8 weeks.)
- Test streaming algorithm for 3D Morse-Smale complex construction. (8 weeks.)
- Implement parallel algorithm for 3D Morse-Smale complex construction. (8 weeks.)
- Test parallel algorithm for 3D Morse-Smale complex construction. (8 weeks.)
- Present results and obtain feedback from stakeholder. (1 day.)
- Choose best of streaming and parallel strategies and refine software and develop to deployment stage. (4 weeks.)

Feature Tracking

- Develop strategy to handle increase in data-size due to increase in dimension.
  - Develop streaming algorithm for feature tracking. (3 weeks.)
  - Develop parallel algorithm for feature tracking. (3 weeks.)
- Implement streaming algorithm for feature tracking. (4 weeks.)
- Test streaming algorithm for feature tracking. (4 weeks.)
- Implement parallel algorithm for feature tracking. (4 weeks.)
- Test parallel algorithm for feature tracking. (4 weeks.)
- Present results and obtain feedback from stakeholder. (1 day.)
- Choose best of streaming and parallel strategies and refine software and develop to deployment stage.
  - Integrate 3D segmentation and tracking into visualization and interaction software already developed in stage A4. (4 weeks.)
  - Present software to stakeholder and get feedback. (1 day.)

Task 5. High quality visualizations and movies for publication and distribution.

- Iterate with scientists to produce high quality visualizations for publication as figures (2 weeks).
- Plan movies using low quality rendering (2 weeks).
- Render frames for movies using Povray or other high quality rendering package 50-200 hours per movie.

Task Dependencies

- Tasks 2 through 4 depend on Task 1, but can proceed after development of a read/write module for a representative data format.
- Although Tasks 3 and 4 can be performed concurrently, the lessons learned during Task 3 can speed-up development in Task 4. Therefore we set Task 3 to precede Task 4.

Accomplishments

- Task 1: Topology research presentation by Valerio Pascucci and meeting with John Bell and group at LBNL.
- Task 2: In order to facilitate work on later stages, we used existing raw format scalar data and visualized it using in-house visualization code written using OpenGL.
- Task 3: There is some overlap between the feature tracking work in this stage and with similar work for the Sandia combustion analysis requirement. We have already started 2D feature segmentation and tracking for the Sandia requirement and will re-use the interaction framework developed for the LBNL requirement.

Work targets for the next six months.

- Task 1: In progress, to be completed in the next six months.
- Task 2: In progress: download and test data.
- Task 3: We have already started work on this and expect it to complete in an additional 28 weeks. Test evolving 3D Morse-Smale complex code.
- Iterate at least two times with stakeholder to get feedback on quality of segmentation into volume/tube/lines and preliminary tracking results.

Risks and Opportunities

1. **OPPORTUNITY** In Task 3 – Feature tracking – we may be able to reuse in part or completely the code developed for gird combustion data. Likelihood: medium-high. Impact on project: high. Remediation: verify this possibility and adjust plan to speedup deployment.
2. **Transition from research prototype code to robust deployment code can be difficult.** Likelihood: high. Impact on project: low. Remediation: We introduce several testing and feedback steps in each stage to ensure that the stakeholder is satisfied with the software. Moreover, all new development effort will use the best software engineering practices and focus on designing modular software. Affected tasks: 2 through 6.
3. **OPPORTUNITY** In Task 4 we may be able to reuse in part or completely the code developed for gird combustion data. Likelihood: medium. Impact on project: high. Remediation: verify this possibility and adjust plan to speedup deployment.

4. **Resource scalability: Large resource requirements to handle extremely large data-sizes.** Likelihood: high. Impact on project: moderate. Remediation: We have access to several large high-performance computers in LLNL, LBNL/NERSC, and ORNL/LCF that can handle such data-sizes. Affected tasks: 4.
5. **Software does not scale appropriate to increase in data-sizes.** Likelihood: high. Impact on project: high. Remediation: We incorporate scalability as a design requirement in each of our tasks. In particular, we include a re-design of 3D Morse-Smale computation and feature software to handle large data-sizes. Moreover, we have experience in designing streaming algorithms that are very memory-efficient and can be adapted to the problems at hand. Affected tasks: 4.
6. **Change in requirements from stakeholder.** Likelihood: low. Impact on project: moderate. Remediation: We add feedback steps in each stage to identify any change in requirements as soon as possible. We keep our software designs sufficiently general and flexible to quickly add more functionality as and when required. Affected tasks: 2 through 4.

Next 12-24 months.

- Complete Task 3.
- Task 4: Start date: 9/16/07. Expected duration: 70 weeks.

12.5 Fusion

12.5.1 Interactive Visual Analysis of Large Fusion Particle-Based Datasets

VACET Team: Allen Sanderson, Utah (Team Leader); Lee Myers, Utah; Thiago Ize, Utah.

Customer/stakeholder(s): Stephane Ethier, PPPL. The SciDAC Center for Gyrokinetic Particle Simulations of Turbulent Transport in Burning Plasmas (W. Lee - PI).

Stakeholder need(s): The physicists are currently generating simulations that use millions to billions of particles, with each particle containing multiple scalar and vector data. They would like to have the ability to explore the nature of the particle orbits in an interactive manner. It is in impractical to view billions let alone millions of particles at one time and glean any insight. As such, the physicists would like to have tools that allow them cull the particles with a user defined query or other statistical tools. After some initial exploration of the data including some initial success in doing a query based visualization.

Milestones/Deliverables

1. Exploration of glyphs for representing multivariate data and exploration of stream based techniques for visualizing moving particles.
2. Develop infrastructure with in SCIRun that allows for the processing of queries.
3. Utilize the Manta rendering environment for interactive exploration.
4. Develop infrastructure with in SCIRun that allows for interfacing with third party statistical tools.
5. Explore alternative tools, such as parallel coordinates for exploring the data.
6. In conjunction with the SciDAC SMD Center develop a data storage flow that allows for easy access by both physicists and visualization scientists that allows for post simulation analysis and visualization.

Each of the tasks are independent of one another.

Accomplishments

During this reporting period the majority of the time has been in establishing a working relationship and understanding the underlying problem space so that some initial exploration could take place. For instance, one of the major difficulties was in simply understanding the data and its physical meaning. This is because the data is often stored in a form that is not in a fully derived quantity that can be used for visualization and analysis. From this exercise it was discovered that in addition to visualization and analysis needs there also exists a data management need.

Our initial exploration of the data has focused on evaluation of infrastructure needs for doing queries and for visualizing multivariate particles using SCIRun. Current SCIRun usage for visualization has been focused on either single time slices or multiple time slices of a data but not with query based applications or for where coherence between moving object is critical. As part of our initial investigation we have developed and deployed a very rough query tool to scientists at PPPL. Using this tool we have had an early initial success in using the query tool to find particles that are "trapped". A movie of this has been produced as is available on the VACET website[25].

Work Targets for Next Six Months.

1. **Task 1a.** Enhancements to the SCIRun infrastructure for glyph based representations. This infrastructure is needed so that it will allow the mapping of multivariate data to each degree of freedom in the glyph. Currently much of this is hard coded and the number of variables that can be mapped is limited. Explore new glyph based representations.
2. **Task 1b.** Enhancements to the SCIRun infrastructure to create new visualization representations based on multiple times slices of data. When objects are in motion having visual cues help the observer envision the path. However this requires infrastructure in place that allows data at multiple time slices of data be processed at any given point.
3. **Task 2.** Enhancements to the SCIRun infrastructure to allow queries. Query based visualization is a new area of work. The current infrastructure allows for single time slice pass-fail operations which is not acceptable with multiple time slices where coherence is needed.
4. **Task 3.** Expand the capabilities of our interactive renderer, Manta to more fully support glyphs and streamlining particles. By utilizing multiple CPUs for the rendering a more interactive setting can be achieved.

Risks

1. **Task 1a.** SCIRun GUI utilizes Tck/Tk which is cumbersome for complex GUI operations. Likelihood: low Impact: medium - multiple implementations may be required or limited capability Remediation: Utilize other tools for the GUI.
2. **Task 1a.** Not all glyphs explored with be useful. Likelihood: low Impact: multiple glyphs may be required. Remediation: Explore lots of glyphs.
3. **Task 2.** The data flow infrastructure may not handle the query needs. Likelihood: medium Impact: requires a one off solution outside of the normal SCIRun paradigm. Remediation: Create an external tool.
4. **Task 3.** Additional glyphs (primitives) in Manta Liklihood:Low Impact: some glyphs may reduce performance Remediation: none.

Work Objectives Next Six Months

[25] http://www.sci.utah.edu/vacetwiki/index.php/Image:GTC_particles.mpg

- Complete Tasks 1-3.
- **Task 4.** Develop infrastructure with in SCIRun that allows for interfacing with third party statistical tools. Statistical packages such as the R and GGobi packages provide powerful tools that can potentially allow clusters of particles to viewed. This could provide greater insight.
- **Task 6.** After the awarding of the SciDAC fusion projects we will re-examine this need in terms of the current funding.

Work Objectives Next 12-24 Months

Task 5. Explore alternative tools, such as parallel coordinates for exploring the data. We believe that viewing the data using an alternative coordinate system may provide insight not seen with traditional queries and statistical methods.

We fully anticipate that the needs will change as our interaction and deployment of tools to the physicists continues. For instance, the current early work has shown limitations in the way we access data stored as HDF5 files at the same time showing that the manner in which they store the data could be improved. This is necessary for the queries to have fast look up.

12.5.2 Poincare Analysis and Visualization of Magnetohydrodynamics Simulation Data

VACET Team: Xavier Tricoche, Utah; Allen Sanderson, Utah (Team Co-leaders).

Customer/Stakeholder: Scott Kruger, Tech-X, working with the Center for Extended Magnetohydrodynamic Modeling (CEMM) SciDAC science application.

Customer/Stakeholder Need(s). Tools for automatic topological analysis of Poincaré plots in the study and assessment of Tokamak simulation data produced by NIMROD. The Poincaré plots are obtained by integrating field lines along a magnetic field.

Milestones/Deliverables

- **Task 1.** Deliver tool that extract significant topological information from fractal structures starting from discrete data. Additionally, the transient nature of the underlying phenomenon must be properly accounted for.
  - Write new software that adaptively integrate field lines in the magnetic field.
  - Apply edge detection to a scalar field derived from the field line information.
  - Extract X- and O-points using robust numerical search.
  - Reconstruct associated islands through solution of boundary value problem.
  - Measure relevant quantitative information associated with islands (number, volume, safety factor, ...).
  - Track the resulting structures over time using topological consistency to infer underlying continuous transformations.
  - Duration: 12 months. start date: October 1, 2006. completion: September 30, 2007.
- **Task 2.** Reliably handle the chaotic behavior exhibited by this type of data and convey it in the form of visual uncertainty in the resulting depiction.
  - Use topological structures to drive the integration of field lines in the direct vicinity of the island boundaries and X-points.
  - Apply to the resulting field lines coherence measure to quantify chaos.
  - Visualize this information in the form of a fuzzy color map superimposed on the topology.
  - Duration: 9 months. start date: October 1, 2006. completion July 31, 2007.

- **Task 3.** Integrate tool directly in simulation code to guarantee full accuracy and optimal leverage of existing infrastructure.
  - Link Poincaré analysis code written in C++ to Nimrod Fortran code.
  - Duration: 2 months. start date: April 15, 2007. completion: June 15, 2007.

Task Dependencies

- Task 2 depends on Task 1 since chaos and uncertainty will be identified and quantified with respect to the know topological structure.
- Task 1 will benefit from Task 2 to assess the reliability of the computed measures.
- Task 3 may be concurrent with Task 1; both correspond to a non-visual data processing task.

Accomplishments

Task 1.

- Some subtasks completed that form a prototype analysis software that provides satisfactory results on test maps.
- Application to a broad range of NIMROD data with various set of properties is part of ongoing work.
- Delivered presentation at the University of Kaiserslautern: Characterizing the Topology of an Hamiltonian System Exhibiting Chaos: From the Standard Map to the Tokamak. Corresponding publication is work in progress.

Task 2.

- First three subtasks have been developed algorithmically. Their application to Tokamak data is the object of ongoing work.
- Technique described previously to quantify coherence and chaos has been the object of two recent publications:
  - C. Garth, G.-S. Li, X. Tricoche, C.D. Hansen. Interactive Visualization of Coherent Structures in Transient Flows. Topology-based Methods in Visualization, Mathematics + Visualization, Springer, accepted for publication, 2007. (discusses the visualization aspect of that type of information and how to use it in complement of other visualization techniques)
  - C. Garth, F. Gerhardt, X. Tricoche, H. Hagen. Efficient Computation and Visualization of Coherent Structures in Fluid Flow Applications. Submitted to IEEE Visualization 2007. (further explores visual exploration aspects for arbitrary flows (e.g. fluid or magnetic) in the 3D and 4D settings and introduces a novel framework to efficiently compute high-quality measures of coherence vs. chaos on large datasets)

Task 3. Work is just now beginning.

Work Targets for Next Six Months

- **Task 1.** Application of code developed for subtasks 1 through 4. Expected completion: June 30, 2007.
- **Task 1.** Subtask 5: measure of relevant quantitative properties. Start date: June 1, 2007. Expected completion: July 31, 2007.

- **Task 1.** Subtask 6: tracking of topological structures over time. Expected completion: June 30, 2007.
- **Task 2.** Subtasks 1-3 will be applied to Tokamak simulation data. Expected completion date: July 31, 2007.
- **Task 3.** Integration of analysis software directly into the simulation code. Expected completion date: June 15, 2007.

Risks

- **New method doesn't perform as well as expected on real-world computational data.** Likelihood: low – a similar approach has already proven successful in the analysis of such data. Impact: medium – some of the underlying assumptions of the method would need to be modified, in particular the idea that a fixed scale is sufficient to identify regions of interest in which X- and O-points are subsequently sought. A solution would consist of implementing an adaptive resolution approach similar to the one developed for the IEEE Visualization 2007 submission listed above.
- **Unable to compute quantitative information associated with islands in a reliable way.** Likelihood: low – computation appears straightforward and the reliability of its assessment will benefit from overall methodology developed for Task 2. Impact: low – increased computational load should address the problem effectively.
- **Unable to reconnect topological structures observed in consecutive time steps because of fast transformations that would exceed the time resolution of the simulation.** Likelihood: medium – this might very much be case specific and depends on the assumptions of the simulation and the quality of the run. Impact: low – a pure semantic analysis of the discrete evolution of the structural contents over time (e.g. we have n_i island chains with safety factors $s1_i$, ..., sn_i at time t_i) should already provide some very valuable information about the evolution of the confinement.
- **Identification of chaotic vs. coherent regions could prove more challenging than anticipated.** Likelihood: medium – the structure of the considered data is intrinsically challenging and practical results are needed to get a true sense of how this method will perform. Impact: medium – different metrics could be used to better discriminate between coherence and chaos in Tokamak data. Very high resolution analysis with the downside of much higher computational cost could also become necessary to address these problems.
- **Difficulties integrating analysis code with simulation code.** Being unable to effectively connect the analysis code to the existing simulation code (requires mixed programming between C++ and Fortran 90), or poor performance of the analysis as a consequence of a suboptimal interface. Likelihood: low – there is significant experience at the SCI Institute in tying C++ code to Scientific Computing code, which is traditionally written in Fortran. Impact: medium – if the simulation code cannot be leveraged to support part of the computational needs of the analysis code, its internal data representation and field line integration algorithms will need to be replicated in a C++ environment, which is likely to take several months to complete. We have done something similar to this at an early stage of this research, though.
- **The topological information is not reliable enough to support the intended type of diagnosis.** This risk is for tasks in the 12-24 month time period. Likelihood: medium – this is definitely an ambitious research and deployment endeavor. Impact: medium – successful completion will probably require the time frame of 12 to 24 months to fully explore this approach and make it meet the needs of this simulation code.

Work Targets Next 12-24 Months

The goal of the upcoming 12 to 24 months will be to integrate the analysis and visualization methods described in the simulation loop so as to provide efficient and accurate diagnosis for large scale computations. Specifically we ambition with the help of Scott Kruger to be able to leverage our framework to diagnose early on deteriorations of the magnetic confinement that will lead to unstable configurations. Expected start date: October 1, 2007. Expected completion date: September 30, 2009.

12.6 Mathematics

12.6.1 Production-Quality AMR Visual Analysis Infrastructure

VACET Team Members

- LBNL: W. Bethel (Temporary Team Leader), G. Weber
- LLNL: H. Childs, B. Whitlock, K. Bonnell
- ORNL: J. Meredith
- UC Davis: Ken Joy
- Related VACET team members: VACET Software Engineering Team (volume rendering work)

Customer/Stakeholder(s)

- Phil Colella, LBNL and the SciDAC Applied Partial Differential Equations Center (APDEC)
- John Bell, Center for Computational Sciences and Engineering (CCSE), LBNL; S. Woosley's Computational Astrophysics Consortium SciDAC Science Application
- Ravi Samteney, PPPL; Center for Extended Magnetohydrodynamics via APDEC

Stakeholder Need(s) Both primary stakeholder teams – APDEC and CCSE – have a need for production-quality AMR visual data analysis infrastructure. Both teams have an in-house tool they have been using and funding over the years. Both teams have expressed the urgent need to get out of the business of developing and maintaining such software infrastructure. Additionally, both teams have expressed the need for new capabilities that lie outside the scope of what is possible with their existing technology.

A more detailed list of specific needs is documented on the VACET twiki[26]. That list contains around twenty different features requested by customers from APDEC and the Astrophysics SAP project. Because they are so numerous and since many will take a significant amount of time to accomplish, our approach is to prioritize these features to optimize the intersection between (1) the stakeholder's assessment of the feature's importance and (2) the degree of difficulty in implementing the feature. With this prioritization, we order our work to achieve impact quickly while continuing work on more difficult and challenging research and engineering problems. The "list of twenty" is itemized below along with labor efforts, stakeholder priority, and status (in some cases where appropriate).

Milestones/Deliverables/Objectives

- **Task 1.** (High priority) Fix slow performance on rectilinear AMR grids. Estimated time: 4 weeks. Status: Completed.

[26]http://www.sci.utah.edu/vacetwiki/index.php/Collab:AMR_Applications

- **Task 2.** (High priority) Visualization spreadsheets capability. Hank has been in discussions with representatives from both groups and has derived a list of features. A design document is posted on the VACET twiki[27]. Estimated time to complete remaining work: 6 weeks.
- **Task 3.** (High Priority) Invoke VisIt from the debugger. The primary work here is to write a VisIt component that connects to VisIt's main state (in the "viewer") and is able to communicate with GDB. The current implementation AMRVis/ChomboVis requires starting up the tool for each new patch. We would like to avoid that here, especially because there have been complaints about startup time for the existing solutions and VisIt not do any better in terms of startup time. Estimated time: five weeks. This work is described in more detail on the VACET twiki[28] as well as in this PMP in Section 12.6.2.
- **Task 4.**(High Priority) Slider for animating through slices. This should be easy, but the slider needs to make sure that each "advancement" corresponds to one logical index, and that index information is not accessible in VisIt's GUI. So the real work here is to add that meta-data information. Estimated time: two weeks.
- **Task 5.** (High Priority) Selecting logical coordinates. This feature maps well to VisIt's IndexSelect operator, but that operator would have to be made AMR-aware (to state what level the logical indices are referring to). The slider needs to make sure that each "advancement" corresponds to one logical index, and that index information is not accessible in VisIt's GUI. The bulk of the work here is to add the needed meta-data information. Estimated time: one week.
- **Task 6.** (High priority) Ensure "new" VisIt runs on platforms important to Stakeholders. This task is expected to be without substantial complications since VisIt has been ported to work on essentially all modern parallel platforms. Estimated time: two weeks.
- **Task 7.** (High/medium) Familiar, customizable interfaces. The approach is to use a ".visitrc" file so that the UI can be customized to appear as either AMRvis or ChomboVis. Estimate: six weeks.
- **Task 8.** (High) Embedded boundaries. APDEC has said they will not switch entirely to VisIt until there is embedded boundary support in VisIt. ChomboVis has an existing solution. Work here is underway with VACET team members from UCD to generate a rigorous mathematical solution that would eventually find its way into VisIt. That work is described in more detail on the VACET wiki[29] as well as in this PMP (Section 12.6.3).
- **Task 9.** (High) Enhancement: It takes too much work to get an outline of the AMR patches. This isn't as trivial as it sounds, because the data travels through a data flow network, and you have to be able to still draw outlines after contours, slices, etc. Estimate: one week.
- **Task 10.** (Medium) Better volume rendering. This milestone is the subject of a significant amount of VACET work as described on the VACET twiki[30] as well as in this PMP (Section 12.7.1). The preliminary implementation of SLIVR, the new volume rendering library, is expected to be finished within six months (by around October 2007). After that time, a preliminary version can be integrated into VisIt. Later efforts will focus on extending SLIVR to run in parallel/scalable mode on shared- and distributed-memory hardware. At the present time, we are unable to estimate a completion date for this milestone.
- **Task 11.** (Medium) Bug fix: axis labels not correct for the scatter plot and for the elevation operator. (3 days)

[27] http://www.sci.utah.edu/vacetwiki/images/0/00/Visual_Spreadsheets_design_spring2007.pdf
[28] http://www.sci.utah.edu/vacetwiki/index.php/Collab:Meredith_Debugging_With_VisIt
[29] http://www.sci.utah.edu/vacetwiki/index.php/Collab:UCD-Spring2007-Status
[30] http://www.sci.utah.edu/vacetwiki/index.php/Collab:Volume_Rendering_Lib

- **Task 12.** Bug fix: some discontinuities were noted with AMR data attributed to poor handling of ghost data. Estimate: 6 weeks.
- **Task 13.** (Medium) Mapped grids (Curvilinear AMR grids). While VisIt already supports curvilinear grids, including curvilinear AMR grids, the Chombo file format for curvilinear grids is presently in flux. (3 days, once APDEC finalizes their data storage format and interface for curvilinear AMR meshes.
- **Task 14.** (Medium) 3D LIC/Vector field visualization. This feature requires a long lead time due to the need to develop new technology combined with significant engineering challenges. This effort is described in more detail on the VACET twiki[31] as well as in this PMP (Section 12.9.1).
- **Task 15.** (Medium) Face- and edge-centered data. (cost needs to be assessed; 6 months+ for complete solution).
- **Task 16.** (Low) Enhancement: Streamlines do not cross AMR patch boundaries with the desired level of continuity. (1 month)
- **Task 17.** (Low) Multiblock domains grids. (1/31/07 note: there is no current support for this in the Chombo file format and they won't have such support for several months). (2 weeks)
- **Task 18.** (Low) User-supplied callbacks. For both "expressions" (i.e., derived quantities) and "queries" (i.e., data analysis routines). Bell wants to be able to provide his own callbacks that perform expensive physics computations. (cost needs to be assessed; 3 month estimate)
- **Task 19.** (Low) Enhancement: picks. VisIt's pick code has primarily the correct functionality, but it could be made more efficient for their use cases. One specific enhancement is that you can add new variables w/o having to repick. (3 days)

Accomplishments

- Approximately half a dozen meetings with APDEC/CCSE/VACET team members to discuss needs to generate the work list.
- Participation in APDEC Kickoff meeting Oct 18-19, 2006 at LBNL. Hank Childs gave a presentation on VisIt, and Valerio Pascucci gave a presentation on Topological Analysis and Visualization.
- APDEC representation at the VACET AHM in SLC February 2007.
- New volume rendering capability being adapted for use in VisIt. This work, which is ongoing by the VACET software engineering team, aims to deploy high quality, multidimensional transfer function capable volume rendering into VisIt. See the Software Engineering Team Design page for more information.
- Visualization Spreadsheet capability designed, tested and added to VisIt. This feature was deemed critical to success for APDEC conversion and has received positive feedback from the APDEC team. New enhancements fell into three major categories: (1) allowing a plot to modify its attributes from the viewer (which meant that plots must be able to access the "ViewerProxy" interface); (2) allowing a plot to access the full meta-data information about the data it is plotting; (3) allowing a plot to instantiate and manage its own window, even one that is not Qt-based. See Figure 7 for a screenshot of the new capability.

[31] http://www.sci.utah.edu/vacetwiki/index.php/Collab:Weber_SurfaceLIC_Project

- VisIt Performance improvements for AMR. Spent approximately one month re-writing key VisIt algorithms to perform better on rectilinear grids. Slowness in VisIt was noted by the teams of Phil Colella (APDEC) and Bell, and this work was to make VisIt more attractive to their teams. Work consisted of fast-tracking common algorithms to operate on and maintain an implicit mesh form (i.e. rectilinear) instead of converting to an explicit form (i.e. triangle mesh). The algorithms addressed were: 2D pseudocolorings, slicing, and looking at the external faces of 3D grids.
- VisIt 1.6 released in April 2007. The new release contains these new features/capabilities.

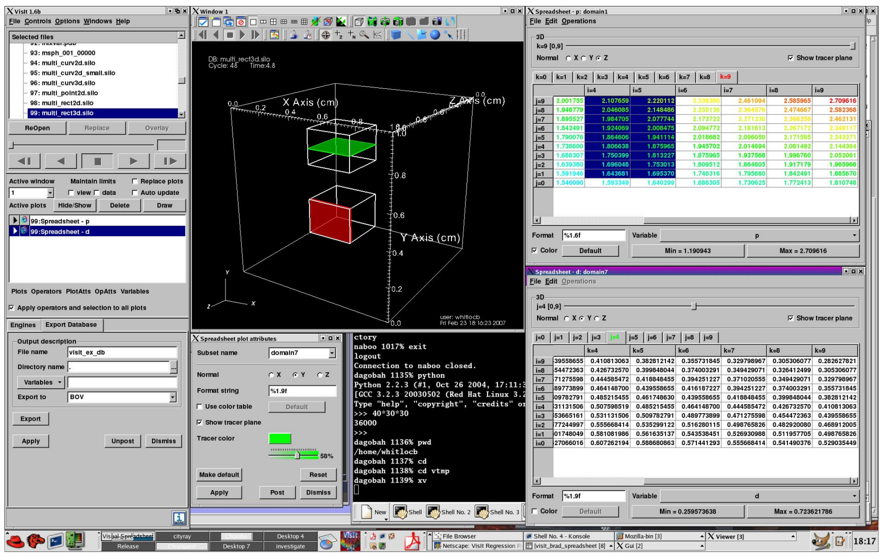

Figure 7: This image is a screenshot of the new AMR-compatible visualization spreadsheets capability in VisIt. This new feature was requested by our stakeholders, developed and released in production form in April 2007.

Work Targets for the Next Six Months

- Tasks 3, 4, 5, 6, 7, 9, 10, 11, 12.
- Preliminary LIC-based vector field visualization in VisIt.
- Migrate VisIt source repository to the SciDAC Outreach Center's GForge server.

Risks

- **Task 3.** Risks associated with this task are described in Section 12.6.2.
- **Task 4.** Possible underestimation of time required; this task will require new infrastructure in VisIt. Likelihood: medium. Impact on project: could require an additional week's worth of effort.
- **Task 6.** Risk: stakeholder requires support for new, unforeseen platforms. Likelihood: medium. Impact: low. The cost estimate for this task takes into account this unforeseen item.
- **Task 6.** Risk: the unforeseen platform has an "unfriendly" operating system (e.g., lightweight microkernels on Crays). Likelihood: low. Impact: medium. Remediation: resolving this issue could require upwards of a month's worth of effort.

- **Task 7.** Stakeholder's vision of successful task completion may not be aligned with current plans. We do not believe this to be true due to a high degree of communication with the stakeholder. Likelihood: low. Impact: low. Remediation: depending upon the degree of difference, implementing new capabilities for the ".visitrc file" should not require much time – a few weeks at most. Remediation: continue discussions with stakeholder.
- **Task 10.** Risk: integration of SLIVR and VisIt is difficult. Likelihood: low – the relevant VACET Software Engineering Team members are pursuing a straightforward strategy for the FY07 implementation plans. Impact: low.
- **Task 10.** Risk: multidimensional transfer editor requires a large investment. Likelihood: high. Impact: medium. Remediation: investigate using SLIVR widgets, which Utah SW team have placed into the SLIVR library; moderate software re-engineering.
- **Task 12.** Risk: solving the coarse-fine boundary issues may lead to extreme "scope creep." Likelihood: medium. Impact: medium – lots of unforeseen coding, possible software design work. Remediation: in the short term, pursue a non coarse-fine solution, with an eye towards a coarse-fine solution in the future.

Work Plans Next 12-24 Months

- Tasks 2, 8, 13, 14, 16, 19.
- Production LIC-based vector-field visualization in VisIt.

12.6.2 Debug With VisIt

Note: this section drills more deeply into one of the milestones listed in Section 12.6.1, namely being able to invoke a visual analysis application, VisIt, from a debugger while running a simulation code. Please refer back to Section 12.6.1 for details about the Stakeholder, VACET team members, and so forth.

Milestones/Deliverables

- **Task 1.** Supplant existing ChomboVis debugging infrastructure with VisIt.
  1. Launch visit using command from debugger.
  2. Support creation of spreadsheet plots in VisIt through the debugger using data from executable.
  3. Support creation of pseudocolor plots in VisIt through the debugger using data from executable.
- **Task 2.** Improve primary usability obstacles.
  1. Improve startup time (e.g. leave VisIt running and amortize startup across all debugging tasks).
- **Task 3.** Further improve usability.
  1. Support new plot types.
  2. Improve integration between running VisIt and running debugger (e.g. new GUI interface).

Dependencies

- Task 1 must occur before Tasks 2 and 3.
- Task 2-1 must occur after Task 1-1, but may be tied to implementation of Tasks 1-2 and 1-3.

- All Task 3 milestones are speculative, and as potentially unnecessary, should occur after all Task 1 and 2 objectives have been met.

Accomplishments

- Milestones for Task 1 (Subtasks 1-3) and milestones for Task 2-1.
  - Expected start date: May 2007.
  - Expected duration: 3 months.
  - Risk: spreadsheet/pseudocolor plots not sufficient for existing functionality. Chance of occurring: low – pseudocolor plot is mature; spreadsheet plot was designed with these customer goals. Impact if it does occur: medium. Remediation: plots would have to be enhanced, but the amount of time required to address deficiencies is likely to be low.
  - Risk: requested data is too large for a single-process VisIt engine Chance of occurring: low – existing mode of operation appears to require on only one patch at a time, which is unlikely to overflow a single-process VisIt engine. Impact if it does occur: medium. Remediation: support for parallel VisIt launching from inside the debugger would have to be supported, but this may only be a simple user interface enhancement.

Next 12-24 Months

Optionally, Task 3 milestones, but only if Task 3 becomes a Stakeholder requirement.

12.6.3 Embedded Boundaries

Team

VACET: John Anderson, UCD; Hank Childs, LLNL; E. Wes Bethel, LBNL; Kenneth I. Joy, UCD (Team Leader)

Non-VACET: Mark A. Duchaineau, LLNL.

Stakeholder(s): Phil Colella/APDEC, LBNL; Mark A. Duchaineau, LLNL; VisIt Users where material boundary calculations are a feature of VisIt and heavily used.

Stakeholder Need(s): In many applications it is necessary to reconstruct or track the boundary surfaces, or "interfaces," between multiple materials that commonly result from multi-fluid Lagrangian-Eulerian hydrodynamics calculations. This problem, where the generated data sets have the characteristic that each cell contains a "fraction" representing the percentage of each material contained in the cell, now arises in a variety of applications, and is frequently called the **embedded boundary problem**. The challenge is to utilize the material fractions in each cell to reconstruct the boundaries between materials. VACET researchers have been developing a number of possible solutions to this problem.

Having a production-quality solution to this problem will allow one of our primary stakeholders and his team (Colella/APDEC) to completely adopt VisIt for their visual data analysis infrastructure. The benefit to their team is a substantial cost savings – they will no longer have to create, update and maintain their own code base for visual data analysis.

Milestones/Deliverables/Objectives

We expect to develop a number of solutions to this problem. We do not expect that any one algorithm will complete solve this problem, but that we must produce a number of solutions that can impact the community. There are several problems that must be solved: (1) the original volume fractions must be preserved, (2) the solution must be scalable, (3) the solution must be

computationally “quick”, and (4) adaptive techniques must be developed that can be tailored to the user’s requirements. We will implement this into VisIt to get the maximum impact with the scientists.

Accomplishments

We have initially developed several solutions to this problem. We are developing a dual-contouring approach that enables adjustment of initial boundaries to conserve material fractions. We have also developed a mathematically-defined guidance surface that enables triangles of varying sizes to be generated by the algorithm and allows interface reconstruction with specified error bounds on the volume fractions. These techniques will be fully developed in the next six months, and prototype solutions will be tested in the VisIt environment.

Project page: http://www.idav.ucdavis.edu/ joy/VACET/EmbeddedBoundary.pdf.

12.7 Software Engineering

12.7.1 SCIRun Library for Volume Rendering (SLIVR)

The primary focus of this project is software engineering to produce a “library version” of the high quality volume rendering technology in SCIRun for use in other areas, notably VisIt. The initial motivation for this work stems from stakeholder comments expressing a desire for better volume rendering in VisIt.

We aim to add 2d transfer function capability to VisIt’s volume rendering infrastructure, similar to that present in SCIRun. Our approach is to separate the volume rendering engine and supporting code from SCIRun and turn that code into a separate library, which we call “SLIVR.”

Another desired outcome for VisIt is to have better compositing, even for 1d transfer functions. VisIt currently does not have shading for its ray casted volume rendering mode, which leads to often poor quality pictures.

There are two broad phases of this project. The first is the software engineering work needed to create a library version of the SCIRun volume rendering technology. Part of this engineering effort includes using the new library in VisIt. The second broad software phase to software engineering to support a parallel use model on both shared and distributed memory architectures, particularly in VisIt.

VACET Team Members: Hank Childs, LLNL (Co-Team Leader); Marty Cole, Utah (Co-Team Leader)

Customer/stakeholder:

The initial motivation for this work stems from comments from some stakeholders requesting high quality volume rendering in VisIt. In the longer term, all VACET stakeholders who want to use volume rendering will benefit from this new capability.

Stakeholder need(s).

High quality volume rendering, multidimensional transfer functions to create more insightful images of multidimensional data.

Subtasks/milestones/deliverables.

Task 1. Identify and relocate core volume rendering code from the SCIRun repository.
Task 2. Remove/recode to reduce the amount of SCIRun infrastructure code dependencies.
Task 3. Establish SLIVR code repository and build infrastructure.
Task 4. Modify a branched version of SCIRun to link against and test the new library.
Task 5. Full regression tests. Merge with the main SCIRun development code branch.

Task 6. Integrate SLIVR with VisIt.
Task 7. Create 2D transfer function editor front end in VisIt.
Task 8. Integrate editor with SLIVR.
Task 9. Develop algorithm for distributed GPU rendering based on existing code.
Task 10. Test and bug-fix the new algorithm.
Task 11. Optimizations and performance improvements: e.g., decrease memory footprint and runtime.
Task 12. Clean up SLIVR interface(s).
Task 13. Developer documentation for using SLIVR.

Task Dependencies

Tasks 1 through 13 are serial in terms of dependency.

Subtasks: Marty is handling the SCIRun side, Hank the VisIt side. They are collaborating on new algorithms in SLIVR. Specifically, Hank is responsible for tasks 6-8, although he may take existing code from Marty for task 7. Marty is responsible for the rest of the tasks, although Hank will likely provide input on 12 and 13 (and potentially others).

Estimated duration:

Project began in November 2006 and is expected to be completed in November 2007.

Accomplishments

Tasks 1 through 4 have been completed during the past six months. Task 5 is in progress.
Task 7: in progress, discussed by VACET SEG team in regular calls. See Figure 8.
Task 9: in discussion.
Tasks 6-13: in progress.

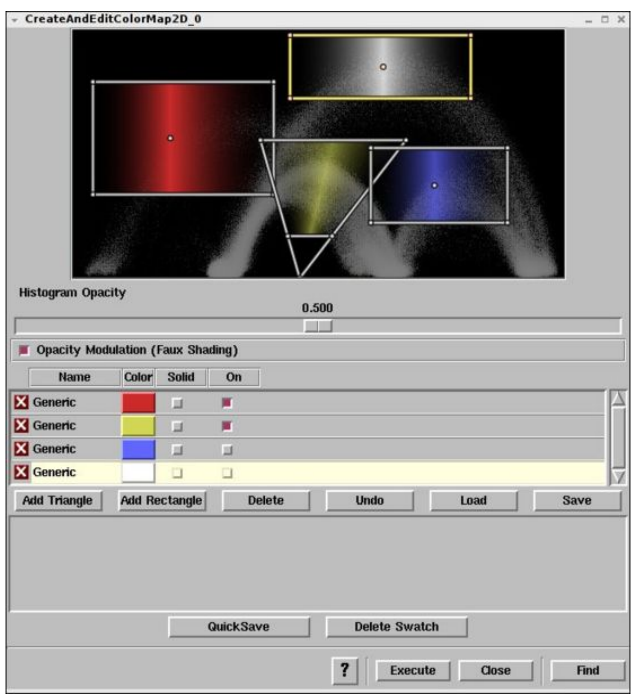

Figure 8: This image shows the SCIRun 2D Transfer Function Editor interface for volume rendering.

Work targets for the next six months.

Work will continue on each of the tasks in sequential order with an estimated completion date of November 2007.

Risks

1. The biggest risk at this point is Task 7. This design my require a number of optimization passes, which may be somewhat open ended time wise. The desire is to have a highly usable interface for 2D transfer function editing. The "worst case scenario" is that the initial interface is not as usable as it could be. The impact is that a less usable interface has a steeper learning curve. Probability: medium. Impact: low. Remediation strategy: to some extent, usability optimizations are somewhat outside the scope of this project. We will remediate this risk by evaluating multiple options within the team then select the one we feel is the best. Evaluation will likely include use with different stakeholder datasets, and some exposure to stakeholders for feedback.

Next 12-24 months. Estimated project completion by November 2007.

12.7.2 VisIt Transition to Subversion

Team:

- **VACET:** Hank Childs, LLNL (Team Lead).
- **Non-Vacet:** The VisIt software team from LLNL..

Stakeholder.

There is no one single stakeholder for this project. It is primarily aimed at improving accessibility to the VisIt code base as well as providing better access/service to all who want to download and use VisIt. Therefore, the stakeholder base is very broad. In addition, the ASC program is a beneficiary, and is providing personnel/effort in support of this project who will work with Childs and others on the VACET team.

Stakeholder Needs

Presently, there is no public access to the VisIt source code. This can be a problem if you want to run VisIt in parallel execution mode – the system must be compiled from source; there is no simple nor straightforward mechanism to provide binaries for parallel use due to the high degree of variability parallel execution environments. Typically, these are very site specific in implementation.

Milestones/Deliverables/Subtasks

- **Task 1.**Establish a repository at NERSC. Convert the existing VisIt repository to SVN. Get existing developers accounts on `svn.nersc.gov`.
- **Task 2.** Make modifications to the code so that the code will build, run, and pass regression tests.
- **Task 3.** Customize Subversion environment to ease developer transition, for example with merging.
- **Task 4.** Get nightly regression testing working using the svn.nersc.gov code base.
- **Task 5.** Team moves away from ClearCase at LLNL to `svn.lbl.gov` at NERSC.

Task Dependencies

- Task 2 depends on Task 1.
- Tasks 3 and 4 depend on Task 2.
- Task 5 depends on Tasks 3 and 4.

Estimated Duration

- Tasks 1 and 2: completed.
- Task 3: 3 days.
- Task 4: 2 days.
- Task 5: 3 days (for training).

Accomplishments

- Tasks 1 and 2: Completed.
- Task 3: in progress.

Next Six Months We expect to have completed all Tasks 1-5.

Risks

- Risk: Customization of Subversion for VisIt is difficult. Likelihood: low. Impact on project: low. Remediation: effort may double. Only Task 3 is affected.
- Risk: additional, unforeseen requirements arise before transition can be completed. Likelihood: medium – Childs has vetted a 9 page document on the conversion that has a 19-point checklist for doing the transition. Only the most prominent 5 bullets appear above as Tasks 1-5. Impact on project: medium. Remediation: The previously mentioned document was the best way to remediate this; it only serves to reduce the risk. Affected task(s): Potential addition of new task.

12.8 Comparative Visualization, Analytics and Analysis Technologies

12.8.1 Multidimensional Visualization and Analytics – Parallel Coordinates

VACET Team: Sean Ahern, ORNL (Team Leader) and Jeremy Meredith, ORNL.

Stakeholder. For this work, there is no one specific stakeholder. Instead, this project focuses on designing and implementing infrastructure that would be applicable across a potentially large number of stakeholder projects.

Stakeholder Needs. Any project that needs to explore large, multidimensional/multivariate datasets would benefit from this work. During R&D, we have been using particle-based datasets produced by the GTC code from the Fusion community.

Milestones/Deliverables

- **Task 1.** Summary information for high number of variables on large data sets.
    1. Implement 2D pairwise binning for arbitrary variable sequences.
    2. Add bin plotting to existing Parallel Axis plot with contrast enhancement.
    3. Add support for many-domain data sets, including in parallel.
    4. Add user interface for runtime selection of parameters.

- **Task 2.**Focus areas of detailed exploration based on 1 or more variable ranges.
    1. Integrate summary information rendering (need A) with existing functionality to render all data points individually.
    2. Integrate user interfaces for the all-data-point ("focus") and summary ("context") plotting.

- **Task 3.** Automated outlier detection (without requiring manual plotting focus).
    1. Add basic outlier detection and plotting.
    2. Update infrastructure for multipass approach needed for parallel/multidomain detection.

- **Task 4.** Make interface and interaction more intuitive and usable.
    1. Migrate visualization window-based widgets to explicit native GUI controls.
    2. Create new window dimensionality supporting more intuitive navigation (e.g. dimension-independent zooming and panning).

Accomplishments

Completed all milestones for Tasks 1 and 2. The new capability has been integrated into the VisIt code base and delivery to the community is imminent. An example of the interface is shown below in Figure 9. Additionally, Figure 10 shows this new data visualization and selection interface being used to select and display particles that are "interesting."

Work Targets Next Six Months

Begin work on Tasks 3 and 4. Expected duration is two to three months.

Work Targets Next 12-24 Months

None.

12.8.2 Representative Particle Field Sampling

VACET Team: Sean Ahern, ORNL (Team Lead); George Ostrouchov, ORNL; Jeremy Meredith, ORNL.

Stakeholder(s): Stephane Ethier, PPPL of the SciDAC Center for Gyrokinetic Particle Simulations of Turbulent Transport in Burning Plasmas (W. Lee - PI).

Stakeholder Need(s): Better particle visualization and analysis. Given the vast number of particles, representative statistical methods for downsampling are needed. The essence of a "representative" sample is to present all "distinct" behaviors along with a measure of how many particles are represented by each behavior.

Milestones/Deliverables

- **Task 1.** Exploration of the problem and working relationships.
- **Task 2.** Write a reader in R for HDF5 GTC particle files.
- **Task 3.** Identify types of useful visualizations that need to scale.
- **Task 4.** Build prototypes of sampling schemes in R.
- **Task 5.** Build prototype methods for evaluating errors due to sampling in R.
- **Task 6.** Embed sampling schemes in SCIRun and VisIt.
- **Task 7.** Embed sampling error schemes in SCIRun and VisIt.

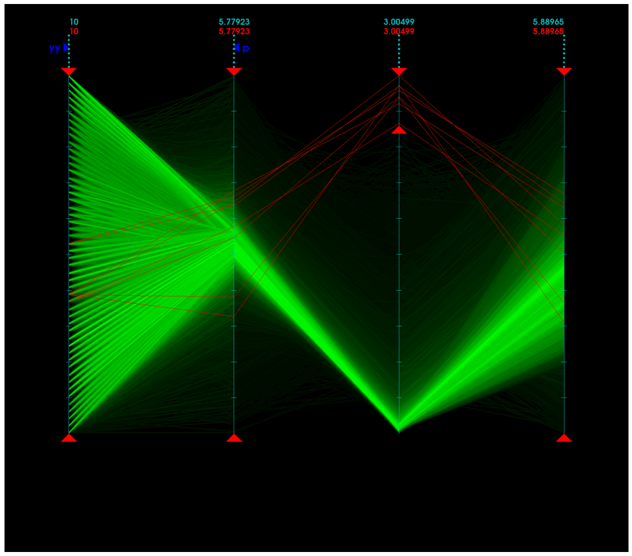

Figure 9: The image shows a Parallel Axis plot in VisIt on different variables of a point cloud interpolated onto a regular grid. The four axes are, in order: the y coordinate, one interpolating function, the magnitude of the gradient of this interpolant, and a second type of interpolating function. The green areas are colored by brightness based on the frequency of the occurrence of data points with the values matching the incident locations on the two adjacent axes. In other words, the areas of greatest brightness show the highest frequency correlations between adjacent variables. Some features apparent here include the spacing of the regular grid in the first axis, the high frequency of very low gradient magnitudes, and the correlation between low gradient magnitudes and non-extreme interpolant values. The red lines show individual (non-binned) data points, and the plotting of these individual values in this image has been restricted to those with only the largest gradient magnitude (the third axis). Thus, the plotting of the individual (red) lines can be used to provide a focus for detailed exploration against the context for the whole data set (in green).

Task Dependencies Tasks are sequential except Tasks 4 and 6 may be performed concurrent with Tasks 5 and 7. Iterating through the tasks will occur over each of several potential different visualization modes.

Accomplishments

- **Task 1** was accomplished through several face-to-face meetings and teleconferences.
- **Task 2** reader is completed and now we are exploring representative data files. The samples need to be representative with respect to the visualization. That is, redundancy needs to be removed for "usual" particles while keeping the extremes. Perspective of what is usual and what is extreme needs to be maintained.
- **Task 3:** Here, we identified visualizations that select particles based on their travel modes within the Tokamak. Various path variability quantifications appear useful in defining the population for sampling.

Work Plans for Next Six Months

For Task 4, we will engage in the following subtasks:

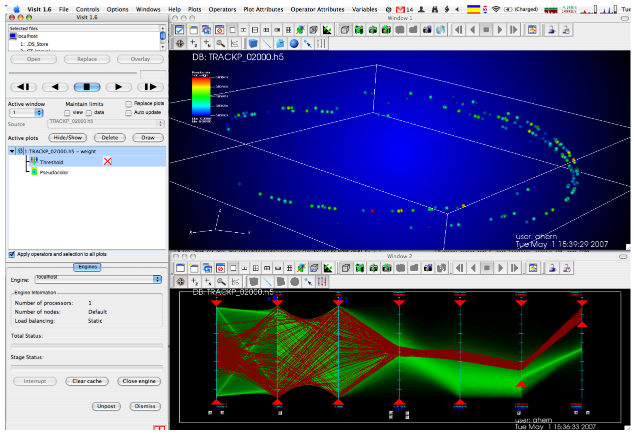

Figure 10: This image shows particle filtering by way of a parallel coordinates plot and selection mechanism in VisIt. The bottom visualization is a parallel coordinates plot of all particles in the timestep. A summary of all data is in green. This parallel coordinates plot is a vehicle for performing an interactive, multi-variate range selection. In this case, the selection is from a range within the “weight” variable to identify particles that are “far from” equilibrium. The data that meet that criterion are drawn using red lines. That data selection is linked to the visualization in the top window. The particles that match the query are drawn as spheres in the upper window, colored by weight.

- Given two or three attributes of interest (existing or derived), select a sample stratified by bins defined in this attribute space.
- Develop a combination of clustering and classification to define distinct clusters in a high-dimensional (3+) attribute space. Select a sample stratified by clusters.

12.8.3 Variable Interactions in Query-Driven Visualization

VACET Team: Luke Gosink, UCD; John Anderson, UCD; E. Wes Bethel, LBNL; Kenneth I. Joy, UCD (Team Leader)

Stakeholder(s): John Bell and Marc Day, LBNL (Combustion Research).

Stakeholder Needs.
Query-driven methods are among the small subset of techniques in which we are able to address petascale data sets. However the challenges of efficient management of the data, and visualization of the results is only part of the problem. With our increasing capacity to generate larger and more complex data sets, there is a need for **scalable methods** that **provide insight** into the trends and behaviors of myriad of variables in a simulation. Our team is developing new methods by which coherent and meaningful visualizations can be constructed that convey information about the interactions between quantities in a query – vastly improving the scientist’s ability to discover interactions between variables in multivariate simulations. We are currently collaborating with John Bell and his team at Lawrence Berkeley National Laboratory in applications of flame-front discovery to illustrate this technology.

Milestones/Deliverables The primary objective of this project is to integrate useful visualization techniques together with query-driven methods. Currently query-driven methods only provide the

user with a “black box” that returns cells that satisfy a query. We wish to develop methods that allow interactions between variables in a query to be visualized.

We expect to integrate these techniques into VisIt and develop modifications for the method according to the needs of scientist stakeholders.

Accomplishments

- **Initial Prototype.** We have developed an initial prototype that allows multiple variables to be integrated into one visualization. Using Cumulative Distribution Functions and Correlation Maps, we can produce visualizations that illustrate the correlations between data against isosurfaces that are defined by the cumulative distribution functions.
- **Paper submission:** L. Gosink, J.C. Anderson, E.W. Bethel, and K.I. Joy, “Variable Interactions in Query-Driven Visualization” submitted to IEEE Visualization 2007.
- New ideas were implemented in VisIt.
- Project page: http://www.idav.ucdavis.edu/ joy/VACET/QueryDrivenVis.pdf.

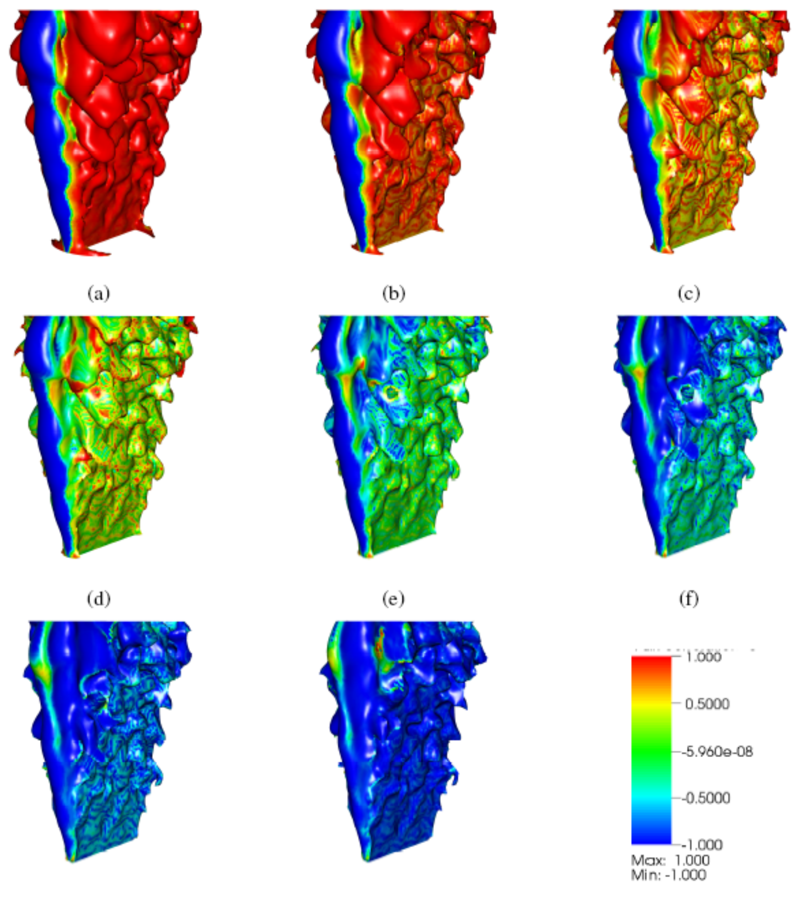

Figure 11: Proceeding from top to bottom and left to right, these images show isosurfaces of increasing values of temperature, which have been rendered through a correlation field constructed from H_2O and C_2H_4. As temperature values increase, we observe a full spectrum of color changes on each isosurface from red to blue indicating that the correlation between H_2O and C_2H_4 is not independent of temperature.

Future Work

In the next 12-24 months, we expect to first integrate these techniques into the work activities of our scientist/stakeholders, allowing us to evaluate their usefulness. We will refine our methods as necessary to produce better techniques, depending on their needs. We also expect to develop new methods of visualization based upon the requirements for visualization of our stakeholders.

12.8.4 Equivalence Class Functions

Team

VACET: Hank Childs, LLNL; George Ostrouchov, ORNL; Sean Ahern, ORNL; Kenneth I. Joy, UC Davis (Team Leader)

Non-VACET: Mark Miller, LLNL; John Clyne, NCAR.

Stakeholder(s). At the present time, this effort focuses on a need that is common across many different science stakeholders. Early versions of this work has been applied to combustion and turbulence research problems. The technique is, however, generally applicable to a very broad range of visual data analysis problems.

Stakeholder Needs.

- Practical tools and algorithms for visualization and analysis of data at extreme scales.
- The ability to quickly and easily generate derived fields reflecting combinations of original data, statistics, or other user-defined operators.

Milestones/Deliverables/Objectives The use of statistical methods in scientific visualization is limited by the lack of scalable algorithms for statistical analysis. Our objective is to work with our National Laboratory stakeholders to develop new scalable statistical techniques that can be used on petascale data sets. The additional analysis capability provided by these methods will greatly increase the data analysis tools available to researchers and scientists.

Accomplishments

VACET Researchers have developed a new class of derived quantities – Equivalence Class Functions (ECFs) – designed to greatly expand the ability of end users to explore and visualize data. These functions are defined over equivalence classes (i.e., groupings) of elements from the original mesh, and produce summary values for the classes as output. ECFs can be used in the visualization process to directly analyze data, or can be used to synthesize new derived quantities on an original mesh. The design of ECFs enable a parallel implementation that allows the use of these techniques on massive data sets that require parallel processing. When integrated and used in concert with standard visualization techniques, ECFs offer a powerful way to manipulate, visualize and analyze scientific data. Formal definition of ECFs and their efficient data-parallel implementation for data sets at the extreme scale will extend the application of many statistical methods to data set sizes they previously could not reach.

- Preliminary implementation of these concepts into VisIt, application to datasets from combustion and turbulence research.
- Publication: M. Miller, H. Childs, J. Clyne, G. Ostrouchov, S. Ahern, and K.I. Joy, "Equivalence Class Functions:
  A New and Versatile Framework for Visualization at the Extreme Scale," IEEE Visualization 2007, submitted.
- Project page: http://www.idav.ucdavis.edu/ joy/VACET/ECF.pdf.

Future Work In the next 12-24 months, we expect to work with our scientist stakeholders to develop new ECF-like functions that can be used to perform statistical analysis of large-scale data sets. We expect to initiate several new projects depending on this analysis.

12.8.5 Comparative Visualization and Analysis

Project Team:
VACET: Hank Childs, LLNL; Sean Ahern, ORNL; Jeremy Meredith, ORNL; Kenneth I. Joy, UC Davis (Team Leader)
Non-VACET: John Clyne, NCAR; Mark Miller, LLNL.

Stakeholder(s): A potentially large audience – those who need the ability to compare datasets with one another. Our deployment vehicle is in VisIt,so this work has the potential to impact a large audience of stakeholders. As a whole, VACET has received requests from virtually all stakeholders for the ability to perform comparative analytics.

Stakeholder Need(s):
VACET researchers have developed a data-level comparative visualization system that is very general in nature – include comparisons in physical space, comparisons in logical space, comparisons across symmetry conditions, and comparisons to analytical functions. Where many previous comparative visualization efforts have focused on A-B comparisons, our system is able to compare many related simulations in a single analysis, allowing for novel visualizations of ensembles of simulations or time-varying data.

Milestones/Deliverables/Objectives Our objectives are to produce new data-level comparison techniques for scientific visualization. The A-B comparison techniques, published in the past, are very limited on the types of data that can be compared. We expect to develop a number of new comparison techniques and integrate them with our VACET software packages.

Accomplishments

- Initial prototype implemented in VisIt. This prototype used for algorithmic testing and evaluation with a variety of different stakeholder datasets.
- H. Childs, S. Ahern, J. Meredith, M. Miller, and K.I. Joy, "Comparative Visualization Using Cross-Mesh Field Evaluations and Derived Quantities," *IEEE Transactions on Visualization and Computer Graphics*, submitted.
- Project Page: http://www.idav.ucdavis.edu/ joy/VACET/ComparativeVis.pdf.

12.8.6 Construction and Visualization of Function Fields

Team:
VACET: John Anderson, UCD; Luke Gosink, UCD; Kenneth I. Joy, UCD (Team Leader).
Non-VACET: Mark A. Duchaineau, LLNL.

Stakeholder(s):
Mark A. Duchaineau, LLNL (initial). As these techniques mature, we expect them to be applicable to radiation transport problems like those that are a part of S. Woosley's Computational Astrophysics Consortium project.

Stakeholder Need(s): Effective techniques for complex and multi-variate visual data analysis.

Objectives/Milestones/Deliverables

Function fields map points in $n-$dimensional Euclidean space to one-dimensional scalar functions. These functions arise in neutron transport applications, hyper-spectral imaging applications, and air-pollution control modeling, among others. Our team is developing new methods for the exploration of such data. In general, they have developed probe-based methods that are used to derive scalar fields that show the similarity structure of function fields. Multiple probe methods allow more information to be seen and provide users with greater visualization flexibility over time and space.

Most of visualization is based upon the display of scalar fields. However, data produced by scientific simulations is multi-variate. We have initiated projects on function fields, where each data point in a mesh is associated with a continuous function. Our objective is to develop, together with our science stakeholders, a number of visualization techniques that allow the exploration of function fields.

Accomplishments and Future Work

Work in this area pre-dates VACET to July 2005. Together with the scientists that study the neutron transport problems (also called “energy-bin problems”), we have developed projects that display function fields. This work has resulted in one initial publication (EuroVis 2007). The multiprobe methods were developed with VACET support.

Our initial work only addressed the problems generated via a function field of scalars. Also within this project are visualizations of function fields, where the functions are more complex (i.e., output scalars). We will address these methods in the next 12-24 months, as well as refine our current methods as needed by our stakeholders.

- Publication: J.C. Anderson, L.J. Gosink, M.A. Duchaineau, and K.I. Joy, “Feature Identification and Extraction in Function Fields,” Proceedings of EuroVis 2007, to appear.
- J.C. Anderson, L.J. Gosink, M.A. Duchaineau, and K.I. Joy, “Exploration of Function Fields using Multiple Probes,” IEEE Visualization 2007, submitted.
- Project page: http://www.idav.ucdavis.edu/ joy/VACET/FunctionFields.pdf.

12.9 Visualization Technologies

12.9.1 3D LIC: Vector Field Visualization

This project focuses on developing and deploying a technique known as Line Integral Convolution (LIC), a well-known approach for vector field visualization. Prior to work on this project, neither of our deployment applications (VisIt, SCIRun) have this capability for visualizing vector fields. Our aim is to provide stakeholders with better vector field visualization capabilities in the production visual data analysis infrastructure.

In addition to production deployment, we aim to achieve several objectives that are novel in nature. First, our design and implementation will be modular in nature so that the LIC field, which is essentially an image in 1, 2 or 3 dimensions that represents a steady-state flow field, can be used to texture any arbitrary geometric primitives. Second, as a texture, rendering performance is accelerated by the hardware texture mapping capability ubiquitous on GPUs. Third, our design and implementation approach aims to provide support for use with AMR data, something that is completely novel.

VACET Team Members:

Gunther Weber, LBNL (Team Leader); Louis Feng, UCD.

Customer/stakeholder:

This new capability was first requested by John Bell early in the needs assessment process.

Stakeholder need(s).

State-of-the-art vector field visualization capability, particularly for use with AMR-based datasets. The stakeholder requested the ability to see a vector field of one variable, e.g., momentum, mapped onto an isosurface of another variable, e.g., temperature.

Subtasks/milestones/deliverables.

Task 1. Implement 3D LIC for a single, uniform resolution mesh.
Task 2. Add infrastructure to VisIt for combining 3D LIC with contour plots (e.g., isosurfaces).
Task 3. Add mapper to VisIt that renders isosurfaces with the LIC texture.
Task 4. Test by resampling AMR grids to a single resolution.
Task 5. Extend to full AMR hierarchies.

Task Dependencies

Tasks 1 and 3 are independent. Task 4 requires completion of Tasks 1 through 3. Task 5 requires completion of Tasks 1 through 3.

Estimated duration:

Tasks 1 through 3 are in progress with an estimated completion date of the end of July 2007.
Task 4 is expected to be completed by the end of August 2007.
Task 5 is expected to be completed by the end of October 2007.

Accomplishments

During this period, the work has transitioned from being an exploratory effort to receiving full attention of the VACET team. Work in the past six months has focused on software engineering, debugging and work on Tasks 2 and 3, including a VisIt mapper that achieves hardware/software OpenGL rendering, including scalable rendering. This work has required close interaction with this team, the VisIt development team and the VACET Software Engineering team to set up the build environment at LBNL and for VisIt consulting.

Risks

1. **Task 1.** No known risks.
2. **Task 2 – VisIt Infrastructure.** The infrastructure required for SurfaceLIC breaks other VisIt functionality (operators/filters). Impact on project: low-medium. Likelihood: medium – there is a high possibility that the data access required for the SurfaceLIC operator interferes with some VisIt operations; the likelihood that this affects desired user-cases is low.
3. **Task 3.** No known risks.
4. **Task 4.** No known risks.
5. **Task 5.** A single texture may not be sufficient for desired surface LIC accuracy Likelihood: medium-high. Potential impact: medium to high. Workarounds: Use high-resolution textures and software renderer or (better) explore possibility of AMR-like texture hierarchy.

Next Six Months.

Complete Tasks 1 through 5.

Next 12-24 months.

Solidify code base to production quality and integrate into VisIt for a production release.

In the longer term, we may wish to create a library version of this capability for use elsewhere. This target may not be practical due to the deep level of platform-specific code needed for an implementation.

12.9.2 Unsteady Vector Field Visualization

Our research has led to the creation of a state of the art technique for the interactive visualization of unsteady flows on geometries of arbitrary complexity and topology. For that purpose we employ a dedicated atlas-based parameterization scheme that lends itself to visualization procedures involving global particle advection over the considered surface. The corresponding computation is mapped to a flat parameter space which is encoded on the texture memory of the graphics hardware to achieve interactive performance. Thanks to these critical features, our framework permits to easily extend most existing dense flow visualization methods to curved surfaces in 3D.

VACET Team: Guo-Shi Li, Utah; Xavier Tricoche, Utah; Chuck Hansen, Utah(Team Leader).

Stakeholder: Any scientific application within the scope of VACET where large 3D vector fields can be analyzed through their restriction to relevant surfaces embedded in the volume. Among them we find Combustion and Fusion.

Milestones/deliverables

- **Task 1.** Parameterize polygonal surfaces of moderate complexity in a way that supports particle advection in computational space.
- **Task 2.** Map corresponding data representation to the GPU.
- **Task 3.** Integrate existing dense flow visualization methods in the framework.
- **Task 4.** Extend the framework to surfaces of arbitrary complexity by improving mesh segmentation and parameterization aspects of the algorithm.

Start date: October 1, 2006.
Expected completion date: September 30, 2007.

Accomplishments

- Tasks 1 through 3 are completed.
- Paper submission: G.-S. Li, X. Tricoche, D. Weiskopf, C. Hansen, Flow Charts: Visualization of Vector Fields on Arbitrary Surfaces. Submitted to IEEE Visualization 2007.

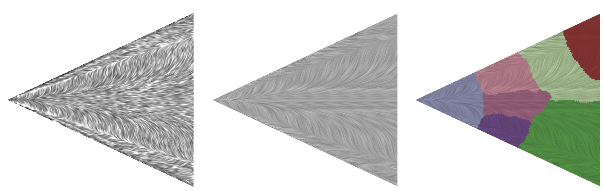

Figure 12: Above figure shows for a delta wing the mesh segmentation used to create the Flow Charts (right image) along with the seamless textures that it generates. Left image corresponds to a dense texture- based flow visualization method called UFAC (Unsteady Flow Advection Convolution, Weiskopf et al., IEEE Visualization 2003), while the middle image corresponds to GPUFLIC (GPU-based Unsteady Flow LIC, Li et al., Eurovis 2006).

Future Plans

- Continue work on Task 4.
- Introduce analysis methods (topology, flow coherence) in the same framework to permit an integrated representation / analysis kind of interactive visualization.

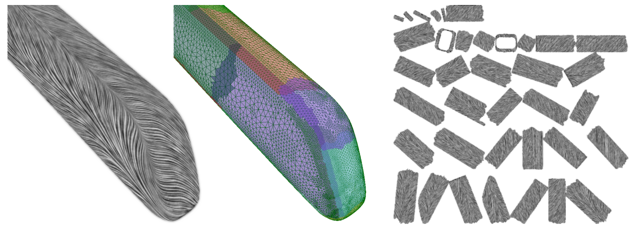

Figure 13: Above image shows similar results for an high-speed train. Left image corresponds to a GPUFLIC representation of the shear stress vector field on the surface of the train. Middle image shows a color coding of the segmentation superimposed on the mesh. Right image shows the packing of the individual pieces of the mesh in texture memory.

12.9.3 Implicit Surfaces and GPU-based Volume Rendering

VACET Team: Chuck Hansen, Utah (Team Leader), Aaron Knoll and Josh Stratton, Utah.

Stakeholder(s) and Needs. These projects focus on fundamental, yet widely applicable technologies and represent more forward-looking aspects of the VACET portfolio. The results of these efforts would be applicable to virtually all VACET stakeholders desiring better/faster single or multi-field volume rendering capability.

Accomplishments

Fall/Spring 06/07: Aaron Knoll has been investigating implicit surface reconstruction in the context of surface reconstruction. This could be used in the ray-tracer. He has developed a SIMD-based interval arithmetic library for use on SSE (CPU) or G80 (GPU) platforms.

Work Plans Next Six Months

Josh Straton will work on developing novel methods for multi-field volume visualization. The plan for Summer 2007 is to re-write the core SCIRun volume rendering and the associated SLIVR library (see Section 12.7.1. Currently, it is register-combiner code; he will generalize it to GLSL.

12.9.4 Parallel, Out-of-Core Visualization Using Hybrid Resources

Team:

VACET: Brian Budge, UCD; Kenneth I. Joy, UCD (Team Leader)

Non-VACET: Tony Bernardin, UCD; Shubhabrata Sengupta, UCD; John Owens, UCD.

Stakeholder(s): This project is in collaboration with the SciDAC Institute for Ultrascale Visualization.

Stakeholder Needs(s): VACET researchers, in collaboration with researchers from the Institute for Ultrascale Visualization are working to develop a new system which enables the rendering of very complex, out-of-core scenes with unbiased global illumination. The system facilitates the rendering of hundreds of millions of triangles, gigabytes of texture, and complex shading by forcing high levels of coherency. A hybrid processing paradigm is created, using both CPUs and GPUs, to exploit the data-parallel nature of rendering. The system has an out-of-core data management base layer designed to support a broad variety of rendering algorithms – including volume rendering for visualization purposes.

Objectives/Milestones/Deliverables

New software/system methods for high-quality visualization are required to process the petascale data sets developed by scientists at the National Laboratories. The objectives of this research effort is explore the feasibility of such an approach.

Accomplishments

Our initial prototype system has been developed. Testing is proceeding on the software infrastructure. A publication detailing this initial prototype has been submitted.

Project page: http://www.idav.ucdavis.edu/ joy/VACET/Parallel.pdf.

Future Work

In the next 12-24 months, we expect to fully develop and understand this problem and have new methods integrated into VisIt. We will work extensively with our stakeholders to define the variants of this problem that must be solved for their uses and initiate new projects accordingly.

13 Risks

Our team has performed an evaluation of risk at both the Center level, as well as on a per project basis. We address Center-level risks below, while a risk analysis for each project has been performed and reported on a per-project basis in Section 12.

13.1 Principal Investigator

Scenario: One of the PIs leave the project. Likelihood: very low. Impact: high. Remediation: Due to the high degree of communication within the VACET center and, in particular, among the members of Executive Committee, the appropriate remediation is to have a backup PI from a lab and one from a University. Bethel and Johnson fulfill the roles of lead and backup PI, respectively.

13.2 Chief Software Architect

Scenario: The CSA leaves the project. Likelihood: very low. Impact: high. Remediation: we have designated Marty Cole (Utah) as the backup CSA. Both the CSA and backup CSA participate in monthly coordination meetings of the VACET software engineering team.

13.3 Current Budget Levels

Scenario: VACET's budget remains fixed at its present level for the entire five years. Likelihood: medium. Impact: Medium-High. When we engineered our proposal, it consisted of several broad but interlocking/interdependent technology areas that are absolutely required to achieve the objective of "PetaScale visualization." If any of them are eliminated, there is significant risk to VACET achieving its objectives. We felt it is in the best interest of our science stakeholders that we retain critical mass in some areas in order to maximize likelihood of impact. The alternative is to uniformly reduce effort across all technology areas. Technically speaking, that approach is viewed as being an even higher-risk – risk is spread to a greater number of project focus areas, thereby increasing overall project risk. Remediation tactics: (1) augment the VACET effort with SAP funding in the out years to pay for one-on-one stakeholder interactions, thereby freeing up some VACET "core funding" to be brought to bear on areas eliminated by the budget reduction; (2) lobby DOE to reinstate our funding based upon initial early successes with science stakeholders;

(3) reduce the amount work from the original proposal; (4) look for opportunities to maximize leverage to achieve technical objectives.

Current Impact. The following list provides an overview of what we believe to be the current impact to our project of a 45 percent reduction in budget.

- **Elimiated the External Advisory Board.** This represents an estimated cost savings of about $25K per year – our original intent was to pay travel costs for EAB members to fly to a single location to conduct an annual review.
- **Reduced Effort in Multiresolution and Progressive Visualization and Rendering.** This effort offered promise for achieving dramatic scalability, but would have required a long lead time and a substantial investment, but would have put us into very good position to penetrate the challenges of Petascale and beyond visual data analysis.
- **Fewer customers.** At the present time, we do not have a concerted effort focusing on Accelerator modeling customers. However, the existing SciDAC Accelerator projects are concluding; the new Accelerator projects will begin later this year. At current budget levels, we are concerned about lack of manpower to dedicate to our Accelerator stakeholders when the new round of SciDAC funding begins. Related, we have been approached by several other Science Applications in the areas of Life Sciences, Materials, Chemistry, Groundwater and Climate; we have had to tell those potential customers that we do not have the ability to help them due to insufficient funding.

13.4 Computational Resources

Scenario: VACET, either collectively or in specific critical projects, is unable to obtain sufficient cycles or other computational resources to meet its stakeholders' requirements. Likelihood: low-medium. To date, VACET collaborators have received significant initial allocations from NERSC and ORNL. However, NERSC's interactive analysis platform, `davinci.nersc.gov`, has nowhere near the capabilities required to realistically achieve our stakeholders needs[32]. Impact: medium to high. Remediation: We would request assistance from DOE in obtaining the necessary resources. Failing that, we would prioritize resources in the direction of the highest priority projects and science stakeholders.

13.5 SciDAC Outreach Center

Scenario: The Outreach Center's scope does not include something like a GForge server that supports software downloads, bug tracking, email lists, collaborative content management, etc. Likelihood: medium. Impact: medium – the adverse impact to VACET is that VACET staff time would be required to set up the infrastructure to do software downloads, role-based collaborative document/content editing, and so forth. This effort will reduce the amount of time we have to accomplish our primary mission. Remediation: for software downloads, we could probably use SourceForge, but this approach is not viewed with favor. For collaborative document management and distribution, we may be able to use Google's 'Docs and Spreadsheets" (see `http://docs.google.com`), although this isn't an ideal solution. Recommendation: VACET recommends that the Outreach Center be immediately commissioned and activated according to the scope of its proposal.

[32]davinci.nersc.gov is a 32P SGI Altix with 192GB of RAM, 24TB of scratch disk and no graphics hardware. We expect requiring a machine with on the order of 128 cores or more, 2TB of memory, 4GB/sec in I/O bandwidth, and the ability to achieve rendering rates on the order of 500M tris/second. To the best of our knowledge, there are no plans in place at NERSC to obtain such a platform

May 2007: This risk seems to have been mitigated to a large degree – DOE has funded the SciDAC Outreach Center at a level commensurate need to provide the GForge-type service. The Outreach Center has recently hired a staff person to bring this new capability online. We await the formal launch of this new service. When fully operational, we will remove this Risk from our PMP.

14 Performance Measures

Generally speaking the primary objective of *Performance Measures*, is to provide the means or a framework for evaluating the degree to which a given endeavor is meeting its objectives. As such, Performance Measures should be derived directly from a project's stated goals and objectives. For many endeavors, defining performance measures is straightforward and they are exclusively quantitative. Examples could include "Is profitability increasing?" In contrast, VACET's mission statement doesn't lend itself as well to traditional, quantitative and objective performance measures.

Earlier in Section 2, we listed a set of Goals and Objectives. Our performance measures, listed below, stem directly from those goals and are consistent with the overall SciDAC program mission.

It is well accepted that performance management of research activities is much different than that for other types of operational or service endeavors. For example, DOE's "Guidelines for Performance Measurement[33]" suggests that different frameworks for developing organizational performance measures are appropriate for different types of projects. The choice of one framework or approach over another depends upon the type of project: a performance measurement approach appropriate for a construction project would not be appropriate for a research project.

Our approach to performance measurement reflects the diversity in scope and time horizon for the different goals in our project. Some performance measures are straightforward and can be obtained through normal operations. Others reflect accomplishments that are possible only over a long period of time and in the presence of favorable circumstances, many of which are outside our scope of control. This mixture reflects our research portfolio, which ranges from basic to applied, but with more emphasis on applied research aimed at meeting scientific stakeholder data understanding needs.

It should be noted that basic research doesn't lend itself well performance measurement in terms of "impact" since such impact often times doesn't occur until decades later. One such example is how the 1960's breakthrough of deciphering the genetic code has led to the identification of genes linked to illnesses such as breast and colon cancer, Huntington's and Alzheimer's disease, and the inception of gene therapy treatments. These impacts could not have been foreseen during the original research during the 1960s.

Finally, we will evaluate the effectiveness of these performance measures over the lifetime of the project. These are subject to change in future revisions of this Project Management Plan depending upon whether or not they prove effective and useful as performance measures.

14.1 Effective and Productive Stakeholder Relationships

Relationships with our customer/stakeholders is a crucial program element. Here, our aim is to work closely with stakeholder/customers to clearly define their scientific data understanding needs, translate those needs into actionable items, prioritize those needs taking into account both stakeholder input as well as VACET-wide considerations (e.g., limited budget, opportunity to reuse technology and effort across multiple stakeholders, unique opportunity for scientific or program-wide impact, etc.).

[33] http://www.orau.gov/pbm/documents/g1201-5.pdf

14.2 Customer Feedback

Customer/stakeholder evaluation is an important VACET performance measure. Here, customers are those who benefit or use, directly or indirectly, the products of VACET's R&D. Customer evaluation is the opinion of one or more customers about either (1) the extent to which VACET's program directly or indirectly benefits the customer; or (2) the extent to which the R&D is perceived as beneficial to the public.

14.3 Effective Production Deployment of Petascale-Capable Technology

As a CET, one of VACET's missions is to deploy petascale-capable visual data analysis software infrastructure at DOE's open computing facilities. In this context, "deploy" is interpreted broadly to include: (1) install the software at DOE's open computing facilities for use by anyone having access to those facilities; (2) run VACET software to process stakeholder datasets at those facilities; (3) provide assistance to the stakeholder in running VACET software at those facilities; (4) run or provide assistance in running VACET software on stakeholder datasets at a location of the stakeholder's choosing[34].

14.4 Coordinated Interactions with Other SciDAC Technology Providers

We anticipate that there will be many opportunities to "combine forces" with other CETs, Institutes and Partnerships to bring to bear diverse technologies to solve stakeholder data understanding problems. This performance measure aims to reflect: (1) the degree to which VACET looks for such opportunities; (2) where appropriate, is able to capitalize upon such opportunities.

Seeking opportunity is expected to be part of our ongoing operations as we participate in community-wide meetings, attend workshops and meetings of our SciDAC sibling projects, have staff from other SciDAC projects attend VACET meetings, and so forth. This type of activity can be planned on an ongoing basis.

Capitalizing upon opportunity is a different matter, and cannot be planned in the same way. Capitalizing upon opportunity requires the confluence of several conditions, which may or may not arise at the right time and may or may not be within our control. Such conditions include: a given technology is sufficiently mature to adopt/apply to new applications in conjunction with VACET technologies; a stakeholder need arises that can be met by multiple technologies from diverse sources; the project teams from which diverse technologies originate have the labor resources to dedicate to a cross-team project. While we cannot predict when such opportunities will arise, we can plan to seek them out through ongoing operations.

14.5 Technical Excellence – Publications

Publications are an important vehicle for conveying research results to the broader scientific community. In addition, most technical publications undergo a process of peer review. The presence of publications in peer-reviewed forums is an important metric that shows our work is of high quality.

[34]This scenario – run at stakeholder facilities – reflects the reality that in some circumstances, it is more efficient to extract subsets of large datasets and move the subset to the stakeholder facility for more detailed analysis. In this case, VACET technology is still being applied to solve stakeholder problems, which meets the larger SciDAC program goal of using computational infrastructure to conduct scientific research.

14.6 Outreach, Education and Service

Over the course of our project, we will also be engaged in several different forms out activities that fall into the broad category of "Outreach and Education." These activities include invited talks/presentations at workshops, conferences and so forth; participation in program-wide meetings, like the SciDAC 2007 program-wide meeting in Boston MA; workshops and tutorials.

A related set of activities falls under the general banner of "Service." These include things like: service as expert on a review committee, organizing and/or participating in planning workshops, service as a Program Committee member, organizer, or a reviewer for technical journals and conferences.

14.7 Train the Next Generation of Visualization and Analytics Leaders

An important aspect of VACET operations is providing the opportunity for junior staff to gain experience in leadership roles to prepare them to serve as future leaders.

15 Communication

15.1 VACET Internal Communication Vehicles

Since the VACET team is distributed both geographically and institutionally, we have come to rely heavily on three different forms of internal communication: email, telephone and wiki.

We have established two different email lists: one for the entire VACET team, the other for the VACET EC. These lists are hosted at LBNL and are quite active. Since we are using *Mailman* to provide email list service, we also benefit from having access to browsable archives of email traffic on both lists.

We have established a standing, toll-free conference call line hosted at LBNL for use by team members. The objective is to make it as easy as possible for team members to quickly arrange conference calls. This project-wide service has proven very useful in ad-hoc meetings to address time-critical specific subjects. Examples of such meetings include: organization of a VACET workshop at the 2007 SciDAC program meeting; several meetings aimed at coordinating a VACET-wide response for Science Application Partnership proposals that are part of the Spring 2007 Fusion SciDAC call for proposals.

We have created a project-wide wiki that is hosted at SCI. The wiki has both public and authentication-required content areas. Its primary role so far is as a vehicle for posting and sharing documents, collaborative document development, individual stakeholder needs assessment and proposed work tasks, per-project management plans, and so forth. Related, SCI hosts a project-wide Subversion server that has proven very useful for collaborative, revision-control document and code development.

15.2 VACET Internal Communication Channels

The VACET management has adopted a position of encouraging frequent communication amongst all team members. This project-wide policy has proven to be highly beneficial in terms of achieving: (1) coordination of project-wide activities; (2) helping each team member to feel an active part of a larger, vibrant effort by knowing what everyone else is working on, project objectives, and so forth; (3) fostering an overall sense of unrestricted and open communication amongst all team members.

The VACET Executive Committee holds bi-weekly conference calls. The purpose of these calls is to discuss project management issues, including but not limited to: (1) stakeholder relations; (2)

defining and prioritizing VACET technical objectives; (3) inter-institution technical interactions. Minutes of these calls are emailed to the entire team. In addition to the regularly scheduled bi-weekly calls, the EC periodically holds out-of-cycle calls to discuss time critical issues when needed.

The VACET Software Engineering Team holds monthly conference calls. Minutes of the calls are posted on the VACET wiki.

In addition to regular calls, these groups periodically hold ad-hoc meetings on an as-needed basis to address time-critical matters.

We have held two VACET-wide, all-hands meetings since the beginning of operations in October 2006. The first all-hands meeting was on 1 November 2006 in conjunction with IEEE Visualization 2006 in Baltimore, MD. The second was held January 31 and February 1, 2007 in Salt Lake City, Utah. The second meeting included representatives from our stakeholder population. More information about the second meeting is located on the VACET wiki[35].

15.3 Communication with Other SciDAC Projects

As indicated in previous sections, VACET team members have been very active in reaching out to stakeholder projects. Such outreach includes email, phone calls, and presence at their project meetings. Based upon the early success of such outreach, we expect such frequent and productive communication to continue in the future.

VACET was well represented at the February 2007 SciDAC project-wide meeting (The "Meet and Greet") in Atlanta, GA. VACET will also be well represented at the upcoming SciDAC program meeting in Boston at the end of June 2007.

As one of the guiding principles for VACET operations is free and open communication, we have a large amount of "surface area" with many different SciDAC projects: each VACET team member is an "ambassador" for the project. So far, this approach has been hugely successful in fostering broad communication with many different current and potential future stakeholders.

15.4 External Communication

Our strategy for outreach and communication to "the world" includes a broad array of activities: web, press, technical publications, and technical presentations.

VACET established a website at `www.vacet.org`. The website, while evolving, is intended to provide a comprehensive view of our Center's activities to a wide audience (science stakeholders, DOE program office, general public).

Press releases provide a valuable way to reach out to a potentially large audience that spans academia, government and industry. One example of such a press release is the November 3, 2006 HPCWire interview with VACET team member Bethel. These types of activities help not only VACET, but also SciDAC and DOE's technology portfolio as a whole.

VACET team members have made presentations in many forums since commencing operations. These are listed in Section 10.9.

VACET team members have a strong record of technical publications. We expect this trend to continue in the future, thereby maintaining VACET's leadership role in the visualization community. These are listed in Section 10.8.

[35] http://www.sci.utah.edu/vacetwiki/index.php/2007_All_Hands_Meeting

References

[1] ASC VIEWS. `http://www.llnl.gov/asci/views`. Advanced Scienfic Computing VIEWS Program.

[2] Scientific Discovery through Advanced Computing. `http://www.osti.gov/scidac/SciDAC.pdf`. Office of Science U.S. Department of Energy.

[3] R. Armstrong, D. Gannon, A. Geist, K. Keahey, S. Kohn, L. McInnes, S. Parker, and B. Smolinski. Toward a common component architecture for high-performance scientific computing. In *Proceedings of the 8th IEEE International Symposium on High Performance Distributed Computation*, Aug. 1999.

[4] F. Bertrand, R. Bramley, K. Damevski, D. Bernholdt, J. Kohl, J. Larson, and A. Sussman. Data redistribution and remote method invocation in parallel component architectures. In *Proceedings of The 19th International Parallel and Distributed Processing Symposium*, 2005. (Accepted, Best Paper Award).

[5] H. Childs, E. Brugger, K. Bonnell, J. Meredith, M. Miller, B. Whitlock, and N. Max. A contract based system for large data visualization. In *Proceedings of IEEE Visualization 2005*, 2005.

[6] H. Childs and M. Miller. Beyond meat grinders: An analysis framework addressing the scale and complexity of large data sets (to appear). In *Proceedings of The 10th International Conference on High-Performance Computing*, 2006.

[7] K. Damevski and S. Parker. Parallel remote method invocation and m-by-n data redistribution. In *Proceedings of the 4th Los Alamos Computer Science Institute Symposium*, page (published on CD), 2003.

[8] L. M. Hwa, M. A. Duchaineau, and K. I. Joy. Adaptive 4-8 texture hierarchies. In *Proceeding of IEEE Visualization 2004*, pages 219–226, 2004.

[9] C. Johnson. Top scientific visualization research problems. *IEEE Computer Graphics and Applications: Visualization Viewpoints*, 24(4):13–17, July/August 2004.

[10] C. Johnson, M. Berzins, L. Zhukov, and R. Coffey. SCIRun: Applications to atmospheric diffusion using unstructured meshes. In M. Baines, editor, *Numerical Methods for Fluid Dynamics VI*, pages 111–122. Oxford University Press, 1998.

[11] C. Johnson and S. Parker. Applications in computational medicine using SCIRun: A computational steering programming environment. In H. Meuer, editor, *Supercomputer '95*, pages 2–19. Springer-Verlag, 1995.

[12] C. Johnson, S. Parker, and D. Weinstein. Large-scale computational science applications using the SCIRun problem solving environment. In *Proceedings of The International Supercomputer Conference 2000*, 2000.

[13] J. Kniss, S. Premoze, M. Ikits, A. E. Lefohn, C. Hansen, and E. Praun. Gaussian transfer functions for multi-field volume visualization. In *IEEE Visualization 2003*, pages 497–504, October 2003.

[14] P. Lindstrom. Out-of-core simplification of large polygonal models. In K. Akeley, editor, *Proceedings of SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, pages 259–262, New Orleans, Louisiana, July 2000. ACM Press / ACM SIGGRAPH / Addison Wesley Longman.

[15] M. Miller, C. Hansen, and C. Johnson. Simulation steering with SCIRun in a distributed memory environment. In E. E. B. Kagstrom, J. Dongarra and J. Wasniewski, editors, *Applied Parallel Computing, 4th International Workshop, PARA'98*, volume 1541 of *Lecture Notes in Computer Science*, pages 366–376. Springer-Verlag, Berlin, 1998.

[16] M. Miller, C. Moulding, J. Dongarra, and C. Johnson. Grid-enabling problem solving environments: A case study of SCIRun and netsolv. In *Proceedings of The 5th International Conference on High-Performance Computing*, 2001 Advanced Simulation Technologies Conference, pages 98–103. Society for Modeling and Simulation International, April 2001.

[17] OMG. Corba Component Model, visited 1-11-2000. http://www.omg.org/cgi-bin/doc?orbos/97-06-12.

[18] S. Parker, D. Beazley, and C. Johnson. Computational steering software systems and strategies. *IEEE Computational Science and Engineering*, 4(4):50–59, 1997.

[19] S. Parker and C. Johnson. SCIRun: A scientific programming environment for computational steering. In *Supercomputing '95*. IEEE Press, 1995.

[20] S. Parker, M. Miller, C. Hansen, C. Johnson, and P.-P. Sloan. An integrated problem solving environment: The SCIRun computational steering system. In H. El-Rewini, editor, *31st Hawaii International Conference on System Sciences (HICSS-31)*, volume VII, pages 147–156. IEEE Computer Society, January 1998.

[21] S. Parker, D. Weinstein, and C. Johnson. The SCIRun computational steering software system. In E. Arge, A. Bruaset, and H. Langtangen, editors, *Modern Software Tools in Scientific Computing*, pages 1–40. Birkhauser Press, Boston, 1997.

[22] S. Parker, K. Zhang, K. Damevski, and C. Johnson. *Integrating Component-Based Scientific Computing Software*, page (accepted). 2005.

[23] W. Schroeder, K. Martin, and W. Lorensen. *The Visualization Toolkit*. Prentice-Hall Inc., 1996.

[24] K. Stockinger, J. Shalf, K. Wu, and E. W. Bethel. Query-driven visualization of large data sets. In *Proceedings of IEEE Visualization 2005*, pages 167–174, 2005.

[25] K. Zhang, K. Damevski, V. Venkatachalapathy, and S. Parker. SCIRun2: A CCA framework for high performance computing. In *Proceedings of The 9th International Workshop on High-Level Parallel Programming Models and Supportive Environments*, April 2004.